

大数据白皮书

（2020年）

中国信息通信研究院

2020年12月

版权声明

前　言

即将过去的 2020 年是历史发展进程中极不平凡的一年。世界正经历百年未有之大变局，特别是突如其来的新冠肺炎疫情为各行各业带来了前所未有的挑战。然而，危机之中，数字化技术驱动的技术和产业变革仍加速发展，大数据技术、产业和应用逆势而上，数据的作用在助力疫情防控和复工复产中大放异彩，“数据驱动”的价值更加深入人心。

今年 4 月，中共中央、国务院发布《关于构建更加完善的要素市场化配置体制机制的意见》，将“数据”与土地、劳动力、资本、技术并列，作为新的生产要素，并提出“加快培育数据要素市场”。5 月 18 日，中央在《关于新时代加快完善社会主义市场经济体制的意见》中进一步提出加快培育发展数据要素市场。数据要素市场化配置上升为国家战略，将对未来经济社会发展产生深远影响。

本白皮书是继《大数据白皮书（2014 年）》《大数据白皮书（2016 年）》《大数据白皮书（2018 年）》《大数据白皮书（2019 年）》之后中国信通院第五次就大数据发布白皮书。本白皮书在此前四版的基础上，回顾了去年以来大数据各领域的最新进展，梳理了中美英等国数据战略的最新动向，阐述了大数据技术的最新趋势，分析了我国大数据产业、重点领域应用以及数据治理的热点与现状，并对“十四五”期间大数据的发展趋势进行了展望。

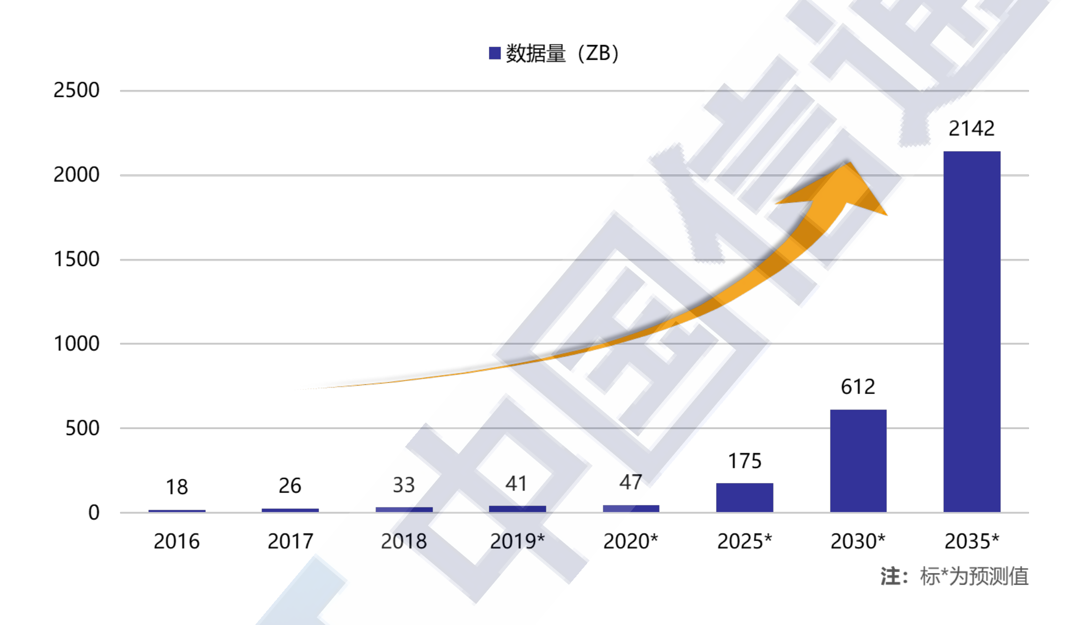
目 录

图 目 录

表 目 录

一、各国的数据战略

当前，全球数据量仍在飞速增长的阶段。根据国际权威机构 Statista 的统计和预测，2020 年全球数据产生量预计达到 47ZB[1]，而到 2035 年，这一数字将达到 2142ZB，全球数据量即将迎来更大规模的爆发。



数据来源： Statista，2020 年 11 月

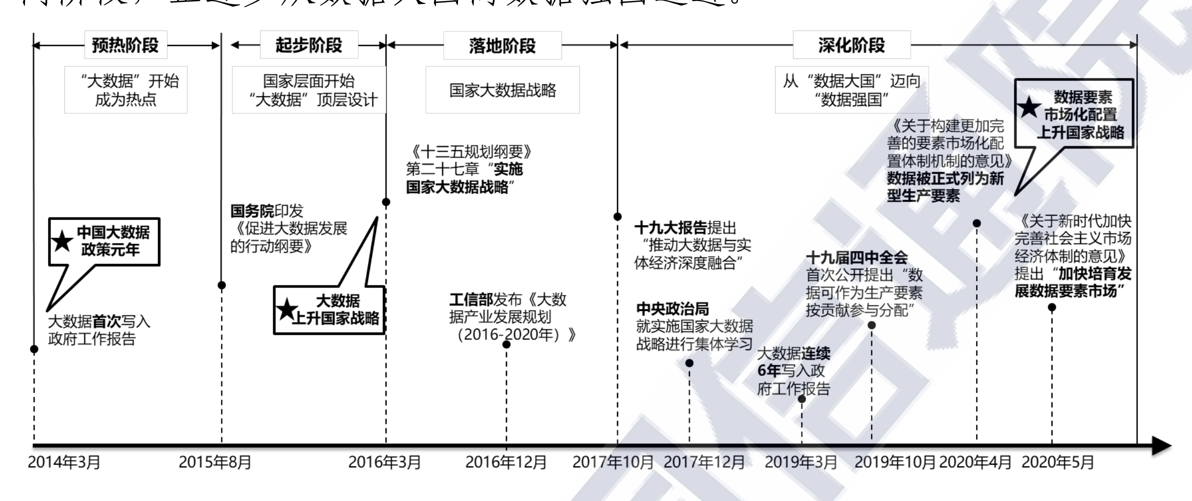
图 1 全球每年产生数据量估算图

随着数字经济在全球加速推进以及 5G、人工智能、物联网等相关技术的快速发展，数据已成为影响全球竞争的关键战略性资源。只有获取和掌握更多的数据资源，才能在新一轮的全球话语权竞争中占据主导地位。2020 年以来，各国数据战略布局步伐加快，本章将梳理我国数据战略的布局历程，并对美欧等各国的数据战略发展重点进行简要叙述。

[1] ZB 为字节计量单位，1ZB=10^{12}GB。

（一）数据要素市场化配置上升为国家战略

自 2014 年以来，我国大数据战略的谋篇布局大致经历了四个不同阶段，正逐步从数据大国向数据强国迈进。



来源：　中国信息通信研究院

图 2　我国数据战略的布局历程

2014 至 2017 年间，国家大数据战略经历了最初的预热、起步后开始落地实施。2014 年 3 月，"大数据"一词首次写入政府工作报告，大数据开始成为国内社会各界的热点。2015 年 8 月印发的《促进大数据发展行动纲要》（国发〔2015〕50 号）对大数据整体发展进行了顶层设计和统筹布局，产业发展开始起步。2016 年 3 月，《十三五规划纲要》正式提出"实施国家大数据战略"，国内大数据产业开始全面、快速发展。

随着国内大数据相关产业体系日渐完善，各类行业融合应用逐步深入，国家大数据战略走向深化阶段。2017 年 10 月，党的十九大报告中提出推动大数据与实体经济深度融合，为大数据产业的未来发展指明方向。12 月，中央政治局就实施国家大数据战略进行了集体学习。2019 年 3 月，政府工作报告第六次提到"大数据"，并且有多项任

务与大数据密切相关。

进入 2020 年，数据正式成为生产要素，战略性地位进一步提升。4 月 9 日，中共中央、国务院发布《关于构建更加完善的要素市场化配置体制机制的意见》，将“数据”与土地、劳动力、资本、技术并称为五种要素，提出“加快培育数据要素市场”。 5 月 18 日，中央在《关于新时代加快完善社会主义市场经济体制的意见》中进一步提出加快培育发展数据要素市场。这标志着**数据要素市场化配置上升为国家战略，将进一步完善我国现代化治理体系，有望对未来经济社会发展产生深远影响。**

我国在国家级政策中将数据定义为“生产要素”，建立在对历史和现实的深入思考之上。人类社会发展的不同时期，都会有相对应的关键性生产要素。这些关键的生产要素都释放了强劲动能，催生了生产技术组织变革，从而拉动了时代快速发展变迁。进入数字社会，数据就成为了这一关键性生产要素。

以史观今，随着人类社会步入数据驱动的数字经济时代，数据要素进一步提升了全要素生产率。在数字社会，数据具有基础性战略资源和关键性生产要素的双重角色。一方面，有价值的数据资源是生产力的重要组成部分，是催生和推动众多数字经济新产业、新业态、新模式发展的基础。另一方面，数据区别于以往生产要素的突出特点是对其他要素资源的乘数作用，可以放大劳动力、资本等要素在社会各行业价值链流转中产生的价值。善用数据生产要素，解放和发展数字化生产力，有助于推动数字经济与实体经济深度融合，实现高质量发展。

从目前来看，作为关键生产要素，大量数据资源还没有得到充分

有效的利用。根据 IDC 和希捷科技的调研预测[2]，随着各行各业企业的数字化转型提速，未来两年，企业数据将以 42.2%的速度保持高速增长，但与此同时，调研结果显示，企业运营中的数据只有 56%能够被及时捕获，而这其中，仅有 57%的数据得到了利用，43%的采集数据并没有被激活。也就是说，仅有 32%的企业数据价值能够被激活。随着数据要素市场培育和建设的步伐加快，数据的有效利用、数据价值的充分释放将成为多方力量共同努力的方向。

（二）各国加快布局探索数据未来发展之路

为了应对信息技术时代在数据方面的发展和挑战，2019 年底以来，美国、欧盟和英国相继出台数据战略，探索未来的数据发展之路。

美国联邦数据战略焦点从“技术”转移到“资源”。自 2012 年以来，美国极力推动大数据领域前沿核心技术的发展和科学工程领域的发明创造，致力打造有活力的数据创新生态。在当前数据成为国家治理重要工具的背景之下，美国政府对于数据的重视程度进一步提升。2019 年 12 月 23 日，美国白宫行政管理和预算办公室(OMB)发布《联邦数据战略与 2020 年行动计划》（以下简称“《联邦数据战略》”），以政府数据治理为主要视角，描述了联邦政府未来十年的数据愿景和 2020 年所要推行的关键行动。

《联邦数据战略》的核心目标是“将数据作为战略资源开发”。《联邦数据战略》确立了 40 项数据管理的具体实践目标，分为三个层面。第一，重视数据并促进共享，如通过数据指导决策、促进各机构间数据流通等；第二，保护数据资源，如保护数据的真实性、完整性和安

[2] 数据来源：希捷《数据新视界》调研，IDC，2020 年。

全性；第三，有效使用数据资源，如增强数据分析能力、促进数据访问形式多样化等。

欧盟数据战略致力于发展数据敏捷型经济体。数据已成为经济社会发展的重要命脉，是新产品和服务衍生的基础。为应对未来发展，欧盟致力于平衡数据流动和广泛使用，希望通过建立单一的数据市场，确保欧洲在未来的数据经济中占据领先地位。2020 年 2 月 19 日，欧盟委员会公布了《欧盟数据战略》，以数字经济发展为主要视角，概述了欧委会在数据方面的核心政策措施及未来五年的投资计划，以助力数字经济发展。

《欧盟数据战略》对欧盟数据发展提出了明确的愿景目标——2030 年欧洲将成为世界上最具吸引力、最安全、最具活力的数据敏捷型经济体。即，在保持高度的隐私、安全和道德标准的前提下，充分发掘数据利用的价值造福经济社会，并确保每个人能从数字红利中受益。为推进欧盟数据一体化和提升欧盟国家的市场主体竞争力，《欧盟数据战略》提出了四大支柱性战略措施：一是构建跨部门治理框架；二是加强数据投入；三是提升数据素养；四是构建数据空间。

英国期待数据战略助力经济复苏。2020 年 9 月 9 日，英国数字、文化、媒体和体育部（DCMS）发布《国家数据战略》，支持英国对数据的使用，设定五项“优先任务”，帮助该国经济从疫情中复苏。这五项任务包括：（1）释放数据的价值；（2）确保促进增长和可信的数据体制；（3）转变政府对数据的使用，以提高效率并改善公共服务；（4）确保数据所依赖的基础架构的安全性和韧性；（5）倡导国际数据流动。

除五项优先任务以外，英国《国家数据战略》还包括多项计划，

如: 到 2021 年, 对 500 名分析师进行公共部门数据科学方面的培训，并设立政府首席数据官，改变政府当前的数据使用方式，从而提高效率并改善公共服务；通过立法提高智慧数据计划的参与度；新建一个 260 万英镑的项目，在支持创新发展的同时致力于解决当前数据共享中存在的障碍等。

除各个国家的数据战略外，国际组织也十分强调数据在全球化发展中的重要性。2020 年 4 月，世界银行呼吁各国政府、相关企业以及学术界共同合作，通过大数据等技术手段应对新冠肺炎疫情所带来的危机。在 2020 年 7 月召开的 G20 数字经济部长会议中，数据流动成为各国部长们重点讨论的议题之一。

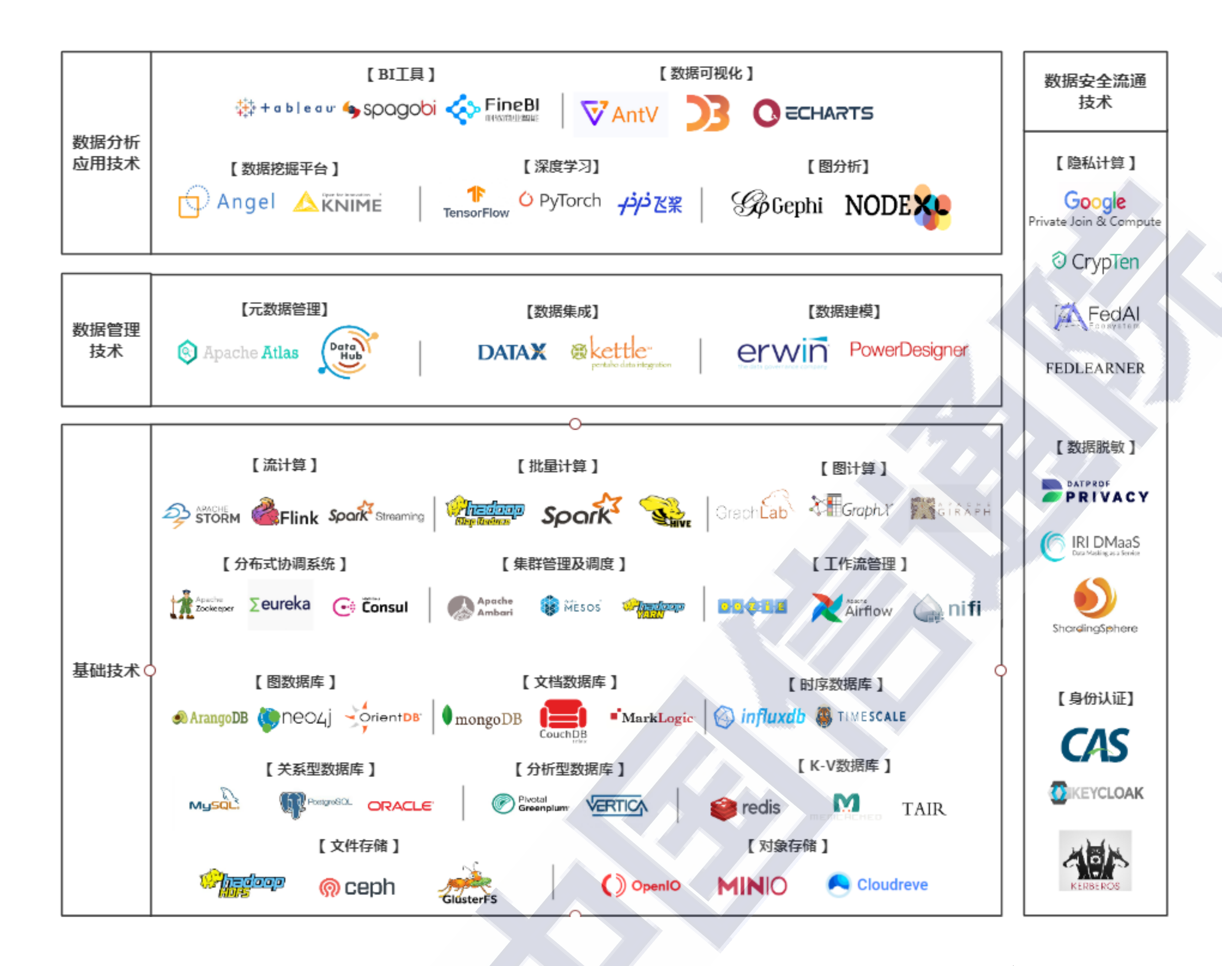
在新一轮的国际经贸规则中，跨境数据流通成为全球双边/多边贸易合作的重要议题。一方面，基于“共同理念”的全球数据同盟体系加速构建，形成了欧盟 GDPR 和 APEC 跨境隐私规则体系（CBPR）两大区域性的数据隐私与保护监管框架，众多国家以二者为蓝本，对本国的数据跨境与数据保护规则进行修订；另一方面，两大框架在国与国、区域与区域之间衍生诸多灵活性的解决方案。2019 年，日韩分别启动与美欧之间的推动跨境数据流动的双边协定，并与欧盟达成充分性保护互认协议。2020 年 3 月，澳大利亚信息专员办公室（OAIC）与新加坡个人数据保护委员会（PDPC）签订关于跨境数据流动的谅解备忘录，加强数据治理方面的合作，促进两国之间的经济一体化；2020 年 6 月，英国宣布脱欧后的未来科技贸易战略，允许英国和某些亚太国家间的数据自由流动，并希望与日本等国达成比其作为欧盟成员国时期更进一步的数据协议。

二、大数据技术发展

近年来，大数据技术的内涵伴随着大数据时代的发展产生了一定的演进和拓展，从基本的面向海量数据的存储、处理、分析等需求的核心技术延展到相关的管理、流通、安全等其他需求的周边技术，逐渐形成了一整套大数据技术体系，成为数据能力建设的基础设施。伴随着技术体系的完善，大数据技术开始向着降低成本、增强安全的方向发展。本章将对当前大数据技术的涵盖范围以及未来发展态势进行介绍。

（一）大数据技术全景解析

大数据技术起源于 2000 年前后互联网的高速发展。伴随着时代背景下数据特征的不断演变以及数据价值释放需求的不断增加，大数据技术已逐步演进针对大数据的多重数据特征，围绕数据存储、处理计算的基础技术，同配套的数据治理、数据分析应用、数据安全流通等助力数据价值释放的周边技术组合起来形成的整套技术生态。**如今，大数据技术已经发展成为覆盖面庞大的技术体系。**图 3 展示了大数据技术体系图谱及相关代表性的开源软件。



来源：中国信息通信研究院

图 3　大数据技术体系及主要开源软件

大数据基础技术为应对大数据时代的多种数据特征而产生。大数据时代，数据量大、数据源异构多样、数据实效性高等特征催生了高效完成海量异构数据存储与计算的技术需求。在这样的需求下，面对迅速而庞大的数据量，传统集中式计算架构出现难以逾越的瓶颈，传统关系型数据库单机的存储及计算性能有限，出现了规模并行化处理（Massively Parallel Processing，MPP）的分布式计算架构；面向海量网页内容及日志等非结构化数据，出现了基于 Apache Hadoop 和 Spark 生态体系的分布式批处理计算框架；面向对于时效性数据进行实时计算反馈的需求，出现了 Apache Storm、Flink 和 Spark Streaming 等分布式流处理计算框架。

数据管理类技术助力提升数据质量与可用性。技术总是随着需求的变化而不断发展提升。在较为基本和急迫的数据存储、计算需求已在一定程度上得到满足后，如何将数据转化为价值成为了下一个最主要需求。最初，企业与组织内部的大量数据因缺乏有效的管理，普遍存在着数据质量低、获取难、整合不易、标准混乱等问题，使得数据后续的使用存在众多障碍。在此情况下，用于数据整合的数据集成技术，以及用于实现一系列数据资产管理职能的数据管理技术随之出现。

数据分析应用技术发掘数据资源的内蕴价值。在拥有充足的存储计算能力以及高质量可用数据的情况下，如何将数据中蕴涵的价值充分挖掘并同相关的具体业务结合以实现数据的增值成为了关键。用以发掘数据价值的数据分析应用技术，包括以 BI（Business Intelligence）工具为代表的简单统计分析与可视化展现技术，及以传统机器学习、基于深度神经网络的深度学习为基础的挖掘分析建模技术纷纷涌现，帮助用户发掘数据价值并进一步将分析结果和模型应用于实际业务场景中。

数据安全流通技术助力安全合规的数据使用及共享。在数据价值的释放初现曙光的同时，数据安全问题也愈加凸显，数据泄露、数据丢失、数据滥用等安全事件层出不穷，对国家、企业和个人用户造成了恶劣影响，如何应对大数据时代下严峻的数据安全威胁，在安全合规的前提下共享及使用数据成为了备受瞩目的问题。访问控制、身份识别、数据加密、数据脱敏等传统数据保护技术正积极向更加适应大数据场景的方向不断发展，同时，侧重于实现安全数据流通的隐私计算技术也成为了热点发展方向。

（二）大数据技术发展趋势

2020 年以来，大数据技术环境发生了一些变化，一些新的技术趋势应运而生，重点呈现出以下几点趋势。

1. 基础技术：控制成本按需索取成为主要理念

大数据技术自诞生以来始终沿袭着基于 Hadoop 或者 MPP 的分布式框架，利用可扩展的特性通过资源的水平扩展来适应更大的数据量和更高的计算需求，并形成了具备存储计算处理分析等能力的完整平台。以往，为了应对网络速度不足、数据在各节点间交换时间较长的问题，大数据分布式框架设计采用存储与计算耦合，使数据在自身存储的节点上完成计算，以降低交互。同时，无论是私有化部署还是云化服务，大数据平台始终以具备数据存储计算处理分析等完整能力的形态提供服务。

存储与计算耦合的自建平台造成了额外成本。实际业务中对于数据存储与计算能力的要求往往是不断变化且各自独立的，使得两类资源的需求配比不可预见且二者到达资源瓶颈的时间无法同步。在存储与计算耦合的情况下，当两者其一出现瓶颈时，资源的横向扩展必然导致存储或是计算能力的冗余，由此必须进行大量的数据迁移才能保证扩展节点的资源得以有效利用，这无疑造成了难以避免的额外成本。同时，以完整产品形式提供服务的大数据平台在应对弹性扩展、功能迭代、成本控制等特性需求时，无论是开发迭代新版本还是集成混搭其他工具，总会引发需求延迟满足、性能持续降低、额外新增成本等其他问题。

存算分离有效控制成本。存算分离是将存储和计算两个数据生命

周期中的关键环节剥离开，形成两个独立的资源集合。两个资源集合之间互不干涉但又通力协作。每个集合内部充分体现资源的规模聚集效应，使得单位资源的成本尽量减少，同时兼具充分的弹性以供横向扩展。当两类资源之一紧缺或富裕时，只需对该类资源进行获取或回收，使用具备特定资源配比的专用节点进行弹性扩展或收缩，即可在资源需求差异化的场景中实现资源的合理配置。

按需索取的处理分析能力服务化概念开始流行。在存算分离理念的基础上，Serverless、云原生等概念的提出进一步助力处理分析等各项能力的服务化。通过存算分离的深入以及容器化等技术的应用，Serverless 概念的落实从简单的计算函数向着更丰富的处理分析能力发展，通过预先实现的形式将特定的数据处理、通用计算、复杂分析能力形成服务，以供按需调用。由此，数据的处理分析等能力摆脱了对于完整平台和工具的需求，大大降低开发周期、节省开发成本，同时服务应用由提供方运维，实行按需付费，消除了复杂的运维过程和相应的成本。

国内外众多厂商深入进行了存算分离和能力服务化的实践。目前，阿里云和华为一众云计算厂商，纷纷提供了基于各自云化大数据平台、分布式数据库产品的存算分离解决方案。其中，阿里云使用自身 EMR+OSS 产品代替原生 Hadoop 存储架构，整体费用成本估算下降 50%；华为则使用了自身 FusionInsight+EC 产品，存储利用率从 33% 提升至 91.6%。在能力服务化方面，国外最为出名的是 Snowflake 公司提出的数据仓库服务化（Data warehouse as a Service，DaaS），将分析能力以云服务的形式在 AWS、Azure 等云平台上提供按次计费的服务，成为云原生数据仓库的代表，并于今年以超过 700 亿美元的市

值 IPO，成为软件企业最大 IPO 案例。在国内则有以阿里云的 AnalyticDB、DLA 为代表的一系列产品提供基于类似思想的服务化的数据处理分析能力。

2. 数据管理：自动化智能化数据管理需求紧迫

数据管理相关的概念和方法论近年备受关注，在大数据浪潮下越来越多的政府、企业等组织开始关注如何管理好、使用好数据，从而使数据能够藉由应用和服务转化为额外价值。

数据管理依赖人工操作带来居高不下的人力成本。数据管理技术包括数据集成、元数据、数据建模、数据标准管理、数据质量管理和数据资产服务，通过汇聚盘点数据和提升数据质量，增强数据的可用性和易用性，进一步释放数据资产的价值。目前以上技术多集成于数据管理平台，作为开展数据管理的统一工具。但是数据管理平台仍自动化、智能化程度低的问题，实际使用中需要人工进行数据建模、数据标准应用、数据剖析等操作。

更加自动化智能化的数据管理平台助力数据管理工作高效进行。在基于机器学习的人工智能不断进步的情况下，将有关技术应用于数据管理平台的各项职能，以减少人力成本提高治理效率成为当下数据管理平台研发者关注的重点。其中数据建模、数据标签、主数据发现、数据标准应用成为几个主要的应用方向。**数据建模方面，**机器学习技术通过识别数据特征，推荐数据主题分类，进一步实现自动化建立概念数据模型，同时，对表间关系的识别将大大降低逆向数据建模的人力成本，便于对数据模型持续更新。**数据标准应用方面，**基于业务含义、数据特征、数据关系等维度的相似度判别，在数据建模时匹配数

据标准，不仅提升了数据标准的应用覆盖面，也减少了数据标准体系的维护成本。**数据剖析方面，**人工智能通过分析问题数据和学习数据质量知识库，提取数据质量评估维度和数据质量稽核规则，并识别关联数据标准，实现自动化的数据质量事前、事中、事后管理。

在数据资产管理概念火热，各项工作备受重视的当下，市场上的数据管理平台产品也在不断演进力争上游。华为、浪潮、阿里云、数梦工场、数澜科技、Datablau 等数据管理平台供应商也在各自的产品中不断更新自动化智能化的数据管理功能。其中华为着重于智能化的数据探索，浪潮关注自动化的标签、主数据识别，阿里云实现了高效的标签识别以及数据去冗，中国系统则聚焦助力数据标准有效落地。

3. 分析应用：图分析需求旺盛引导数据分析新方向

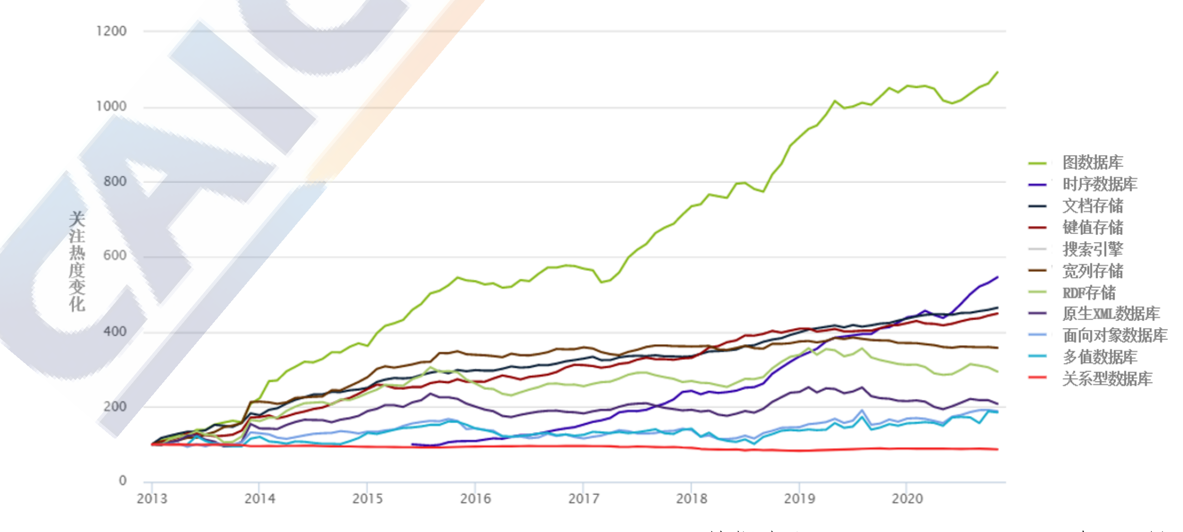
随着深度学习的迅速发展，传统的针对以独立数据集合为对象的分析技术不断成熟。相对的，对于存在关联关系的数据进行关联分析的需求愈加旺盛。关联分析最早始于由上世纪九十年代，由“购物篮分析”问题，即通过从顾客交易列表中发掘其购物行为模式引申而来。早期机器学习领域中也有 Apriori、FP-growth 等经典频繁模式挖掘算法实现对于关联规则的挖掘分析。

传统数据分析方法难以应对图结构数据中关联关系的分析需求。以社交网络、用户行为、网页链接关系等为代表的数据，往往需要通过“图”的形态以最原始、最直观的方式展现其关联性。在图的形式下，自然而然地存在着连通性、中心度、社区关系等一系列内蕴的关联关系，这类依赖于对图结构本身进行挖掘分析的需求难以通过分类、聚类、回归和频繁模式挖掘等传统数据分析方法进行实现，需要能够对

于图结构本身进行存储、计算、分析挖掘的技术合力完成。

专注于图结构数据的图分析技术成为数据分析技术的新方向。图分析是专门针对图结构数据进行关联关系挖掘分析的一类分析技术，在分析技术应用中占据的比重不断上升。与图分析相关的多项技术均成为热点的产品化方向，其中以对图模型数据进行存储和查询的图数据库、对图模型数据应用图分析算法的图计算引擎、对图模型数据进行抽象以研究展示实体间关系的知识图谱三项技术为主。通过组合使用图数据库、图计算引擎和知识图谱，使用者可以对图结构中实体点间存在的未知关系进行探索和发掘，充分获取其中蕴含的依赖图结构的关联关系。

根据 DB-Engines 排名分析，图数据库关注热度在 2013-2020 年间增长了 10 倍，关注度增长排名第一。图数据库、图计算引擎、知识图谱三项热点技术方向正在全球范围内加速产业化。国内阿里云、华为、腾讯、百度等大型云厂商以及部分初创企业均已布局这一技术领域。其中，知识图谱已经开始深入地应用于公安、金融、工业、能源、法律等诸多行业，纷纷落地内部试点应用。



数据来源：DB-Engines，2020 年 11 月

图 4　数据库领域关注度变化态势

4. 安全流通：隐私计算技术稳步发展热度持续上升

除了对数据进行分析挖掘以外，数据的共享及流通是另一个实现数据价值释放的方向。无论是直接对外提供数据查询服务还是与外部数据进行融合分析应用，都是实现数据价值变现的重要方式。在数据安全事件频发的当下，如何在不同组织间进行安全可控的数据流通始终缺乏有效的技术保障。同时，随着相关法律的逐步完善，数据的对外流通面临更加严格的规范限制，合规问题进一步对多个组织间的数据流通产生制约。

基于隐私计算的数据流通技术成为实现数据联合计算的主要思路。在数据合规流通需求旺盛的环境下，隐私计算技术发展火热。作为旨在保护数据本身不对外泄露的前提下实现数据融合的一类信息技术，隐私计算为实现安全合规的数据流通带来了可能。当前，隐私计算技术主要分为多方安全计算和可信硬件两大流派。其中，多方安全计算基于密码学理论，可以实现在无可信第三方情况下安全地进行多方协同计算；可信硬件技术则依据对于安全硬件的信赖，构建一个硬件安全区域，使数据仅在该安全区域内进行计算。在认可密码学或硬件供应商的信任机制的情况下，两类隐私计算技术均能够在数据本身不外泄的前提下实现多组织间数据的联合计算。此外，还有联邦学习、共享学习等通过多种技术手段平衡了安全性和性能的隐私保护技术，也为跨企业机器学习和数据挖掘提供了新的解决思路。

由于解决的问题十分契合数据流通领域的热点命题，近年来隐私计算技术持续稳步发展，各类市场参与者逐渐清晰。一方面，互联网巨头、电信运营公司以及众多大数据公司纷纷布局隐私计算，这类企业自身有很强的数据业务合规需求，也有丰富的数据源、数据业务、

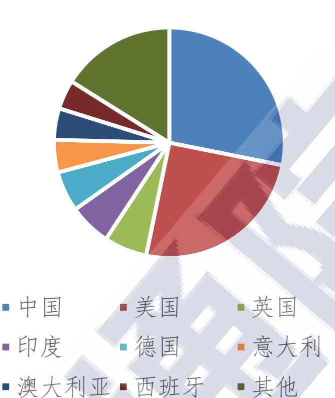
数据交易场景和过硬的研发能力。另一方面，一批专注于隐私计算技术研发应用的初创企业也相继涌现，对外提供算法、算力和技术平台，相关理论技术较为扎实专业。整个隐私计算技术领域开始呈现百花齐放的快速发展态势。

（三）大数据科研创新进展

自 2012 年大数据广泛进入实际应用以来，产业界和学术界在大数据技术与应用方面的研究创新不断取得突破，大数据领域的论文发表数量和专利申请数量快速增长。

在论文发表方面，根据 Web of Science 数据库收录的 SCI 论文数据统计，2012 年至今，全球共发表以“big data”为关键词的大数据领域相关论文 64739 篇。其中，2012-2018 年大数据相关论文发表量持续增长，并在 2018 年达到高峰，2018 年全年共发表相关论文 11453 篇，是 2012 年论文数量的 5.4 倍，七年间的年均增长率约为 32.5%。2019 年论文数量较前一年有所减少，但仍然超过一万篇，2020 年 1-6 月，论文发表数量则为 3915 篇。随着科学研究的不断进展，大数据相关的理论体系将逐渐成熟，未来学术论文成果发表的增长速度或将放缓。从发表论文的国家和地区来看，中国和欧美地区仍是大数据相关学术研究的核心地带。中国和美国是大数据相关论文发表最多的国家，分别发表论文 18216 篇和 16241 篇，占论文总量的 28.14%和 25.09%，遥遥领先于其他国家。英国、印度和德国的论文发表数量占比分别为 6.03%、5.92%和 5.66%，意大利、澳大利亚、西班牙及其他国家的论文数量占比均低于 5%。



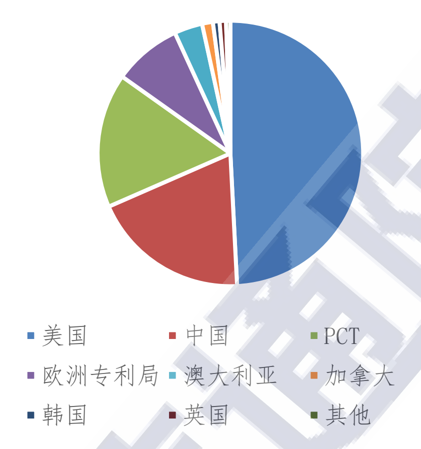


数据来源： Web of Science，2020 年 10 月

图 5　大数据领域论文发表数量及各国占比

在专利申请方面， 根据 WIPO（世界知识产权组织）的 PATENTSCOPE 数据库的统计数据，2012 年至今，全球共申请大数据领域的相关专利 136694 项。2012-2019 年间，大数据技术快速发展，各国和国内外大数据企业对知识产权保护更加重视，专利申请数量始终保持稳定增长，从 2012 年的 9135 项持续增长至 2019 年的 25854 项，年均增速约为 16.0%。2020 年 1-6 月，全球大数据相关专利申请数量则为 10789 项。从受理专利申请的国家和地区来看，大数据相关专利的申请较为集中，93%以上的专利均分布在美国、中国、PCT 组织（专利合作条约体系）和欧洲专利局。美国受理的专利数量位居首位，占比近 1/2，约为 49.19%，IBM、微软、Oracle、EMC 等美国企业的技术优势仍然显著；中国为第二大专利受理国，占比约为 19.25%，一方面，我国大数据发展态势良好，在国际上有重要的市场地位，另一方面在技术创新与突破上，我国与美国仍有较大差距。此外，PCT 组织和欧洲专利局受理的专利数量占比分别为 16.39%和 8.31%；其他国家的占比均在 5%甚至 1%以下。





数据来源： PATENTSCOPE 数据库，2020 年 10 月

图 6　大数据领域专利申请数量及各国占比

从细分领域来看，在大数据领域的科学研究中，针对基于分布式计算技术的数据关联分析或数据挖掘的论文和专利数量明显多于其他领域，可见数据分析挖掘依旧是大数据科研创新的热点方向。

三、大数据产业发展

“十三五”以来，我国大数据蓬勃发展，融合应用不断深化，数字经济量质提升，对经济社会的创新驱动、融合带动作用显著增强。工业和信息化部运行监测协调局发布的数据显示，2019 年我国以云计算、大数据技术为基础的平台类运营技术服务收入 2.2 万亿元，其中，典型云服务和大数据服务收入达 3284 亿元，提供服务的企业达 2977 家[3]，大数据产业发展日益壮大。

本章将结合最新趋势对当下大数据产业的范围和内涵进行了界定，并从产业主体、企业投融资趋势等方面对我国大数据产业发展的态势进行简要分析。

（一）大数据产业生态界定探讨

随着大数据技术不断演进和应用持续深化，以数据为核心的大数据产业生态正在加速构建。尽管自实施国家大数据战略以来，我国的大数据产业发展已取得明显的成绩，但产业相关的基本问题，包含产业内涵、外延与特征等尚未达成共识。厘清大数据产业范畴和体系结构，区分大数据产业核心要素与关联要素，对于推动大数据产业发展起着十分重要的基础作用。在这里，将对大数据产业生态的各类界定方式进行简要叙述。

出于理解视角的差异，大数据产业内涵的界定目前仍有争议。一类观点从产业经济学出发，认为大数据产业是以大数据为出发点和落脚点，通过对自身生产或从外部获取的数据进行挖掘、应用以创造价值的经济活动集合。但也有一类观点认为大数据只是现代信息技术产

[3] 工信部运行局《2019 年软件和信息技术服务业统计年报》，2020 年 10 月

业中的一部分，因为大数据的本质是在互联网、软件、计算机等基础上实现的数据服务，其围绕的数据采集、传输、加工、分析、应用等一系列活动仍包含于现代信息技术产业的范畴之内。

从大数据的价值体现出发，结合以上各类研究机构和学者的观点，我们认为**大数据产业是以数据及数据所蕴含的信息价值为核心生产要素，通过数据技术、数据产品、数据服务等形式，使数据与信息价值在各行业经济活动中得到充分释放的赋能型产业。**

传统的大数据产业定义一般分为核心业态、关联业态、衍生业态三大业态。核心业态是包括从大数据采集到服务、数据交易、数据安全以及相关平台运营建设等围绕数据全生命周期的大数据关键技术与业务；关联业态是以软件、电子信息制造业为代表，包括智能终端、集成电路、软件和服务外包等大数据产业所需的软硬件制造业务；衍生业态是包括工业、农业、金融等各行业的大数据融合应用。

对大数据产业的另一种分类是划分为基础支撑、数据服务和融合应用三层业态。基础支撑层包含网络、存储和计算等硬件设施，资源管理平台以及与数据采集、分析、处理和展示相关的技术和工具；数据服务层是围绕各类应用和市场需求，提供包括数据交易、数据采集与处理、数据分析与可视化、数据安全等辅助性服务；融合应用层则包含了与政府、工业、交通等行业密切相关的应用软件和整体解决方案。

通过对比，不难看出以上两种分类的本质几乎一致。核心业态与数据服务、关联业态与基础支撑、衍生业态与融合应用之间各自相互对应，后一种分类可以看作是围绕前者描述的具体展开。

还有部分观点认为以上两种分类较为笼统，并未将数据资源明确

纳入到大数据产业的相关业态中，而数据资源应是大数据产业链条的起始点，不可忽略。基于这样的观点，有学者提出基于数据价值的实现流程，应将大数据产业分为大数据资源供应、大数据设备提供、大数据技术服务和大数据融合应用四层。这一分类方式是在沿用上述两种分类的基础之上，将基于互联网、物联网等信息技术渠道大量产生并提供数据资源的经济活动单列出来，成为大数据产业链条的第一层。此外，也有观点将大数据产业划分为五个层次，这里暂不展开介绍。

表 1　不同大数据产业分类方式间的对应关系

分类方式	层次名称与定义			
三层分类（一）[4]	**核心业态** 指围绕数据全生命周期、大数据关键技术和大数据核心业务所形成的产业业态，包括大数据的采集、处理、存储、分析、交易、安全、服务和云平台建设运营	**关联业态** 指大数据产业链上下游与大数据核心业态紧密联系的电子信息产业，包括智能终端、集成电路、电子材料和元器件、呼叫服务、电子商务、互联网金融、软件和服务外包等	**衍生业态** 指大数据、互联网+在各行业、各领域的融合应用所产生的业态	//
三层分类（二）[5]	**数据服务** 围绕各类应用和市场需求，提供辅助性服务，包括数据交易、数据采集与处理、数据分析与可视化、数据安全等	**基础支撑** 包括网络、存储和计算等硬件设施．资源管理平台以及与数据采集、分析、处理和展示相关的技术和工具	**融合应用** 包含了与政务、工业、交通等行业紧密相关的应用软件和整体解决方案	//
四层分类[6]	**大数据技术服务业** 贯穿大数据产业链的软件及技术服务提供	**大数据设备提供业** 贯穿大数据产业链的硬件设施提供方，包	**大数据融合应用业** 大数据产业链后端的数据应用方，包	**大数据资源供应业** 大数据产业链前端的数据资源提供

[4] 《中国共产党贵州省第十二次代表大会报告关键词解读》http://szb.gzrbs.com.cn/gzrb/gzrb/rb/20170913/Articel06003JQ.htm

[5] 大数据产业生态联盟，《中国大数据产业生态地图暨中国大数据产业发展白皮书》

[6] 李桥兴，胡雨晴. 大数据产业的属性与分类界定及其模糊识别研究[J]. 科技管理研究, 2020, 040(003):163-173.

	方，包括前端采集、数据清洗、大数据管理分析平台建设、商务智能挖掘等围绕数据提供的相关技术服务及软件研发	括光缆、网络设备、高核能计算机、大数据一体机、集成电路等大数据所需的硬件设备的设计、制造、租赁、批发和零售等	括与互联网、金融、交通、政务等行业的融合，为不同行业提供相应的服务和解决方案以实现经济目标	方，包括移动互联网行业、金融业、电信业、交通运输业等能够产生并拥有大量数据资源的行业

来源：根据公开资料整理，中国信息通信研究院

（二）大数据产业商业模式解读

大数据产业层次划分难以明确统一的原因之一在于各层次之间的企业业务经营存在交叉覆盖。从实践看，以互联网巨头为代表的诸多科技企业在大数据产业上的布局已跨越了多个层次，提供硬件设备、技术软件与应用方案等多类产品与服务。在这里，简要对大数据产业的主要商业模式进行分析。

一是提供数据或技术工具。这类模式以数据资源本身或数据库、各类 Hadoop 商业版本、大数据软硬件结合一体机等技术产品，为客户解决大数据业务链条中的某个环节的对应问题。按照资源的不同分类收费，既可以买断数据资源或技术产品，也可以按需、按月、按年、按量等方式获得付费服务，方便零活。

二是提供独立的数据服务。这类模式主要指为数据资源拥有或使用者提供数据分析、挖掘、可视化等第三方数据服务，如情报挖掘、舆情分析、精准营销、个性化推荐、可视化工具等，以付费工具或产品的形式向客户提供。

三是提供整体化的解决方案。这类模式主要是为缺乏技术能力但需要引入大数据系统支撑企业或组织业务升级转型的用户定制化构建和部署一整套完整的大数据应用系统，并负责运营、维护、升级等。

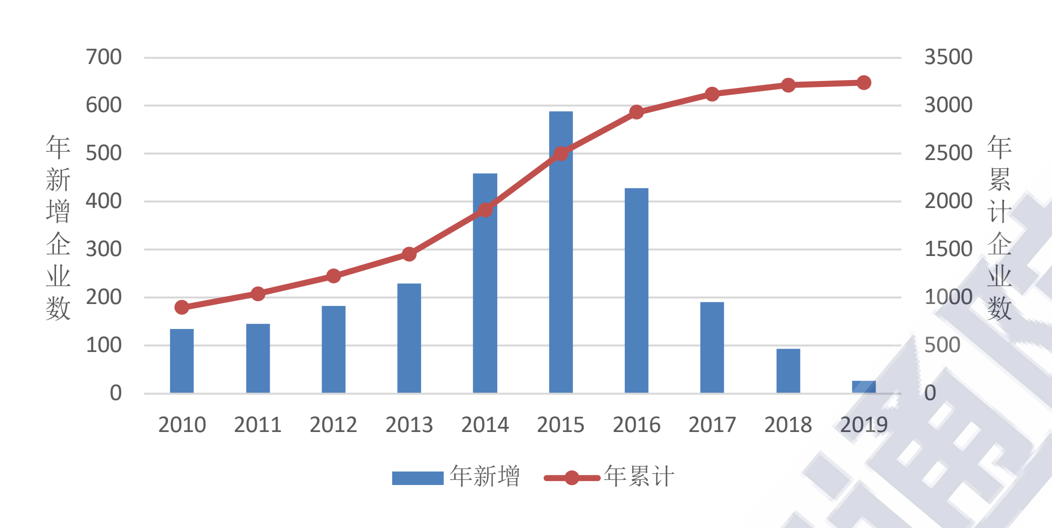
（三）大数据产业主体发展洞察

2020 年，我国大数据产业迎来新的发展机遇期，产业规模日趋成熟。大数据产业主体从“硬”设施向“软”服务转变的态势将更加明显，面向金融、政务、电信、医疗等领域的大数据服务将实现倍增创新，大数据与特定场景的结合度日益深化，应用成熟度和商业化程度将持续升级。在这里，将对我国大数据企业的发展现状进行观察与分析。

1. 大数据企业数量平稳增长

根据中国信息通信研究院监测统计，当前我国活跃的大数据企业共有 3242 家[7]。大数据企业的快速增长阶段出现在 2013 至 2015 年，增长速度在 2015 年达到最高峰。2015 年后，市场日趋成熟，企业新增开始趋于平稳，大数据产业走向成熟。其中企业数量的变化与新政策的出台密不可分。2015 年 8 月国务院颁布《促进大数据发展行动纲要》，大数据由此正式上升为国家发展战略。2016 年工信部印发了《大数据产业发展规划（2016-2020 年）》，推动大数据产业进一步发展。另一方面，新一代信息技术、智慧城市、数字中国等发展战略逐步推动社会经济数字化转型，大数据的产业支撑得到强化，应用范围加速拓展，产业规模实现相应快速增长。

[7] 中国信通院根据公开监测、获取到的企业信息，结合企业介绍、主营业务及服务介绍及相关其他信息，按照从事数据生产、采集、存储、加工、分析、服务等业务类型对企业进行筛选和分类得到。

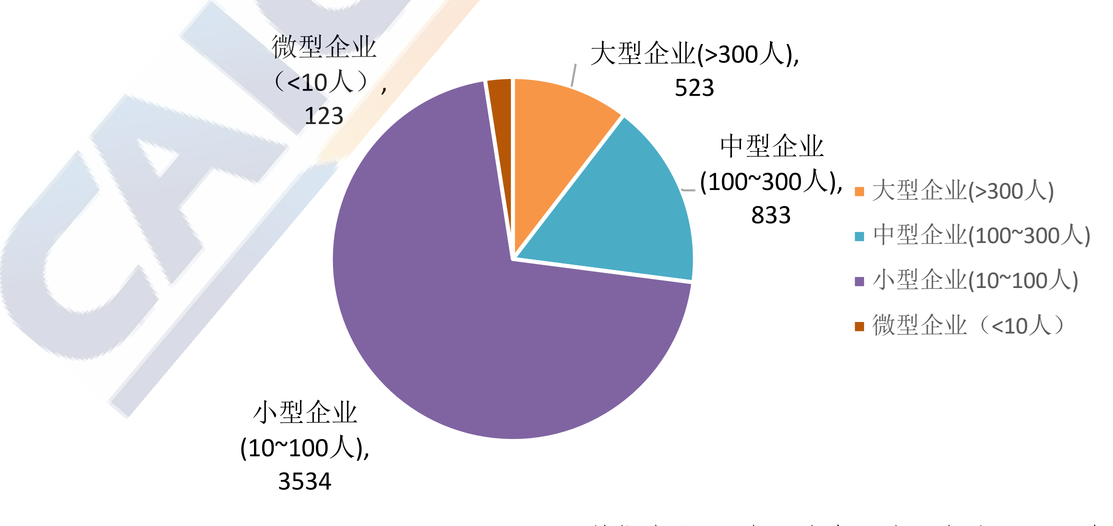


数据来源： 中国信息通信研究院，2020 年 10 月

图 7 大数据企业数量增长统计

2. 10-100 人的小型企业占主导

我国目前大数据领域的企业超 3,000 余家，而超 70%的大数据企业为 10 人至 100 人规模的小型企业，在产业蓬勃向上的发展阶段离不开中小企业发挥其在创新创业中的重要作用。政策上伴随“新基建”成为拉动国内经济发展的新一轮驱动力，大数据中小企业面临的外部市场环境和依托的基础设施也发生重大变化从而影响企业规模分布。



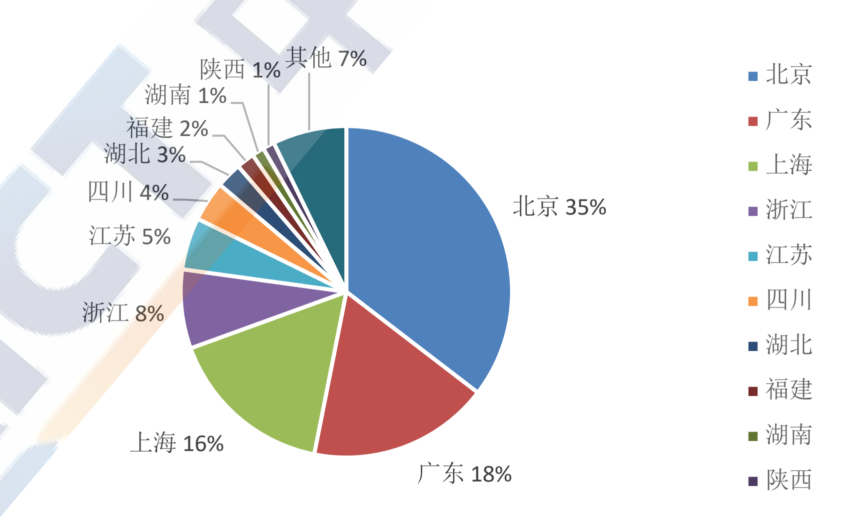
数据来源： 中国信息通信研究院，2020 年 10 月

图 8 大数据企业规模分布

3. 地域分布以北上广为主

根据统计，我国大数据企业主要分布在北京、广东、上海、浙江等经济发达省份。受政策环境，人才创新，资金资源等因素影响，北京大数据产业实力雄厚，大数据企业数量约占全国总数的 35%。依托京津冀大数据综合试验区，天津、石家庄、廊坊、张家口、秦皇岛等地大数据产业蓬勃发展，依靠良好的政策基础、科研实力、地理位置和交通优势，分别形成了大数据平台服务和应用开发、数字智能制造、旅游大数据等创新企业集聚中心，在信息产业领域形成了竞争优势。

2020 年，随着数字经济发展热潮兴起、数字中国建设走向深入，我国大数据产业迎来新的发展机遇期，各区域更重视大数据发展与地区经济结构转型升级的紧密结合，同时各级政府也都更积极地探索数据驱动的政府服务模式创新。

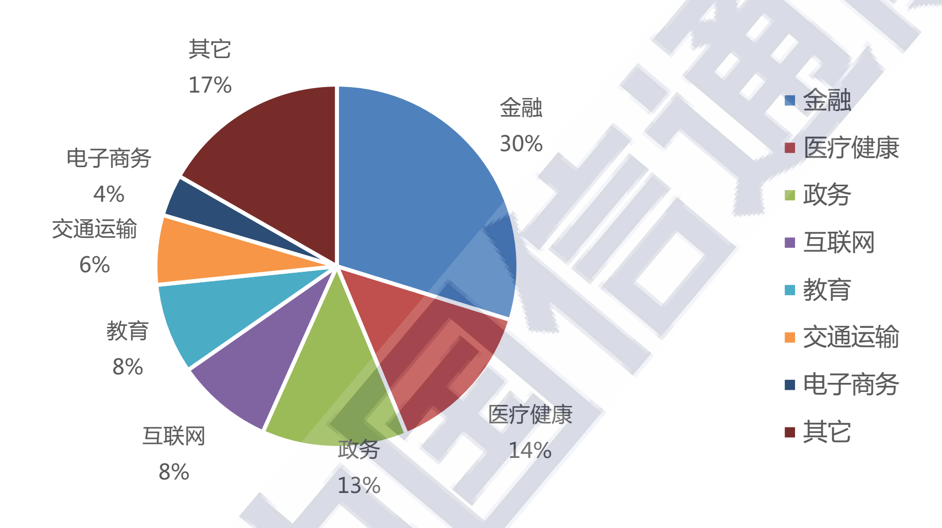


数据来源： 中国信息通信研究院，2020 年 10 月

图 9　大数据企业地域分布

4. 行业应用领域丰富

根据大数据企业的业务标签，将 1404 家涉及行业大数据应用的企业进行了统计整理。图 9 显示出行业大数据应用企业涉及的行业分布。从图中可以看出，金融、医疗健康、政务是大数据行业应用的最主要类型。除此之外依次是互联网、教育、交通运输、电子商务、供应链与物流、农业、工业与制造业、体育文化、环境气象、能源行业。



数据来源： 中国信息通信研究院，2020 年 10 月

图 10 大数据行业应用企业类型分布

（四）大数据企业融资趋势分析

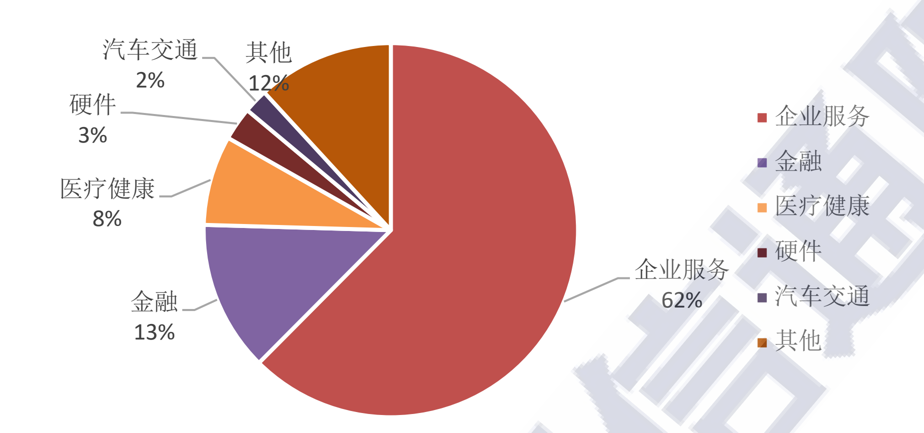
本白皮书还统计了我国大数据企业获得投融资的情况[8]。

1. 企业服务为主要融资领域，金融和医疗前景可期

从融资细分领域分布来看，大数据行业融资企业分布在近 20 个领域，大数据行业迎来历史新机遇，在企业服务、医疗健康、金融等垂直细分领域的大数据应用展现出巨大潜力。大数据产业增量蓝海市场正在逐步打开，截止到 2019 年，企业服务领域的企业获投占比最

[8] 中国信通院根据专业行业投融资数据库进行数据筛选、清洗，并联合大数据企业信息完成整体处理和分析得到。

高为 62%，金融行业次之为 13%，健康医疗为 8%。随着互联网与移动互联网的进一步普及渗透，以及 IT 基础设施的逐步完善，企业服务市场仍将继续扩大。

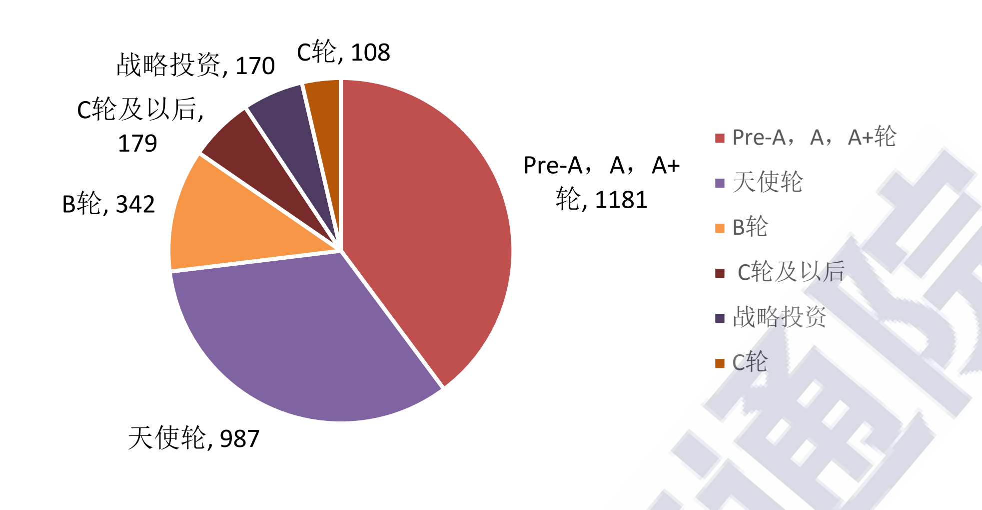


数据来源： 中国信息通信研究院，2020 年 10 月

图 11 大数据企业赛道分布

2. B 轮之后竞争加剧，初创公司经受考验

从融资轮次上看，大数据行业融资主要集中在早期项目，其中天使轮项目占 33%，Pre-A、A、A+轮项目 40%，合计占比 73%。大部分企业仍处于早期投资占主导的阶段，这也可以佐证投资机构对大数据市场依然充满信心。随着行业知识、技术架构等方面日趋成熟，早期初创公司经受过市场的检验，赢得了投资人的信任，融资估值也日趋增高。



数据来源：　中国信息通信研究院

图 12　大数据获投轮次分布

四、大数据行业应用

“十三五”期间，我国大数据融合应用能力不断深化。大数据在工业领域的应用不断深入拓展，驱动网络化协同、个性化定制、智能化生产等新业态新模式快速发展。电信、互联网、金融等重点领域优秀大数据产品和解决方案加速涌现，精准营销、智能推介等应用日益成熟。疫情监测、病毒溯源、资源调配、行程跟踪等大数据创新应用场景快速兴起迭代，在常态化疫情防控中发挥了突出作用。本章将重点介绍大数据在疫情防控、通信、工业、互联网及金融行业应用的最新趋势。

（一）大数据全面助力打赢疫情防控“阻击战”

2020 年开年之际，新型冠状病毒引发的肺炎疫情汹涌而至。经过全国社会上下艰苦卓绝的努力，国内疫情防控阻击战取得了重大的战略成果。回顾此次抗“疫”历程，大数据在疫情监测分析、人员管控、医疗救治、复工复产等各个方面，得到了广泛应用，发挥了巨大作用，

为疫情的防控工作提供了强大支撑。

一是疫情分析展现。疫情期间，很多企业选择从大数据分析和展现入手为政府、公众和企业的疫情防控提供支撑。通过对人员和车辆流动、资源分布、物流运输等信息进行全方位、多角度的实时展示，支撑政府的疫情防范管制。通过疫情信息展示、人流迁徙呈现、舆论监测与评价、民众信息上报与展示等，及时为公众播报疫情信息动态。还有很多企业通过自建或采购疫情分析与展示产品，实现企业内部疫情的有效防控和管理。

二是疫情防控管制。疫情防范类应用通过模型建立、分析挖掘等手段利用位置数据和各类行为数据实现高危人群识别、人员健康追踪、区域监测、市场监管等功能。大量科技企业利用 AI 图像识别、智能外呼、知识图谱、安全多方计算、微服务等多项技术，为社会疫情防控及政府决策提供有力支撑。

三是医疗医治增效。大数据和智能技术在病情诊断、医学研究、医疗辅助等医护工作的相关场景中得到了充分应用。百度研究院开放线性时间算法 LinearFold，提升了新冠病毒 RNA 空间结构的预测速度；浙江省疾控中心基于阿里达摩院研发的 AI 算法上线全基因组检测分析平台，有效缩短疑似病例的基因分析时间，且能精准检测出病毒的变异情况。

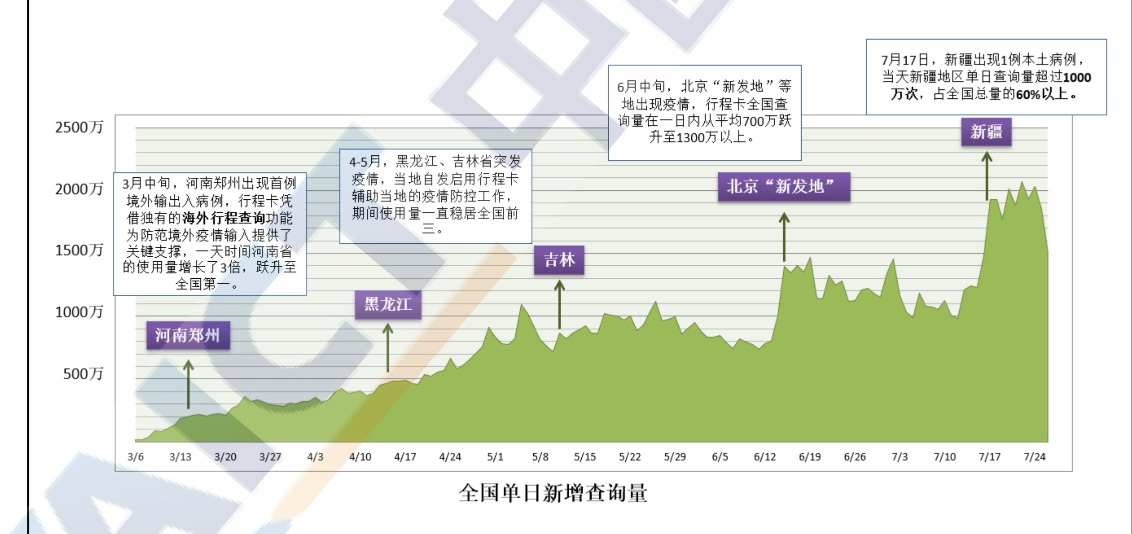
四是生活便民举措。生活服务类应用也是数据驱动疫情防控的重点突破口。诸多互联网企业采用 O2O 服务模式，形成线下活动到线上活动的映射。利用大数据技术实现海量生活数据的采集、分类和存储，为居民提供无接触外送、实时疫情地图、互联网医疗等服务。在便利居民正常生活的同时，确保了各类服务的安全健康。

五是复工复产管理。在复工复产的重要阶段，国务院办公厅电子政务办指导推出的全国一体化政务服务平台疫情防控健康信息码，解决了数据标准不一致的问题，实现了跨省跨地区的疫情服务互联和有序复工复产；百度 Hi 企业智能远程办公平台、阿里“钉钉”、华宇软件等远程办公软件支撑了政企学的异地协同运转。

专栏 1：通信大数据行程卡有效助力疫情防控

2020 年 2 月，在工业和信息化部领导下，中国信通院、中国电信、中国移动、中国联通共同推出“通信大数据行程卡”，并在国务院客户端微信小程序上线，为全国 16 亿手机用户免费提供 14 天内所到地市信息的查询服务。

通信大数据行程卡的技术原理是分析手机“信令数据”，获取用户设备所在位置信息。信令数据的采集、传输和处理过程自动化，有严格的安全隐私保障机制，不与其它个人信息进行匹配，查询结果实时可得且数据全国通用。行程卡 APP2.0 版本还引入了低功耗蓝牙技术（BLE），为用户提供新冠肺炎密切接触者追踪提醒功能。截至 2020 年 11 月，累计查询量已超过 42 亿次。



来源：中国信息通信研究院

图 13　通信大数据行程卡查询量变化情况

（二）需求推动通信大数据价值进一步发挥

通信大数据作为大数据产业的重要组成部分，一直备受产业各方

关注。我国通信大数据发展迅速，应用市场需求不断增长，正处于快速发展期。根据 2020 年 8 月工信部发布的数据，1-7 月，全国电信业务收入累计完成 8027 亿元，大数据、云计算、数据中心等新兴固定业务增长成为电信业务收入增长的第一推动力。数据显示，1-7 月三大运营商大数据相关的业务收入增速达 47%，增势显著。

相对于其他行业，通信大数据具备全面、动态、实时的特点，形成了独特的优势：一是数据资源优势，通信数据规模巨大，基于 16 亿移动电话用户、2 亿固定电话用户、3 亿固定宽带用户，运营商每天可搜集 PB 级的数据；二是基础设施优势，高速的网络带宽、全面的网络覆盖和高效的网络运维等为通信大数据的应用提供了可靠保障。

目前，通信与其他行业间的数据融合成为通信大数据应用的热点方向，公共安全、民生服务、旅游开发、商业推广等众多领域均已有了代表性的实践案例。在通信大数据的应用中，保护个人信息始终是前提。目前主要的应用场景全部都采用经脱敏、泛化后的数据或不针对特定个人的群体统计数据。

专栏 2：电信运营商积极探索跨行业数据融合应用

当前，各大运营商都在积极探索大数据商业化运营，在获得用户授权、保障用户个人隐私的基础上，利用业务运营积累的大数据资源和自身技术能力对外进行商业价值释放。

通信大数据+商业拓展。某运营商利用基站定位和大数据分析技术推出了“商铺选址”业务，面向商业地产、零售连锁、旅游景区等行业，提供商业选址、位置评估、客流分析等服务。

通信大数据+旅游推广。某运营商通过与地方政府旅游局达成战略合作，建设基于通信大数据的旅游产业信息服务平台，可以在为游客提供智慧旅游服务的同时，通过人流统计、客源分析等帮助旅游景区提升服务质量、制定更优的营销策略。

通信大数据+政府治理。某运营商利用基于广域群体的手机信令数据，分析用户群体的活动规律，在不触碰市民个人隐私的前提下，辅助政府部门实时统计辖区内的人口热力分布，预测各地区活跃人口的变化趋势，为政府部门开展人口统计、调控与监测提供有力的技术支撑。

随着 5G 的全面商用、物联网等相关技术的深化拓展和多样化智能终端的逐渐普及，可以预期，通信大数据将有更为广阔的应用空间。“十四五”期间，我国通信大数据将呈现出如下的发展趋势：

一是合规性要求进一步提升。随着我国《网络安全法》《数据安全法》的陆续颁布或征求意见，数据安全与合规成为了通信大数据发展的重要基础。通信大数据涉及用户广、社会影响巨大，如何在确保用户的个人权益的基础上，安全合规地发挥数据价值将成为关键议题。

二是价值释放的路径更加多元。随着多方安全计算等数据流通技术的不断应用，通信大数据有了更多的价值释放途径。例如，可以在运营商环境中布设多方安全计算节点，或通过联邦学习等方式实现“数据不出门”“可用不可见”。

三是应用模式将更加丰富。通信大数据的应用正在向“全行业”模式转变，覆盖智慧政府、城市大脑、金融风控、健康医疗、疫情防控、精准营销等更多领域。

（三）政策铺垫为工业大数据提供发展机遇

随着政策环境的铺垫和工业互联网基础设施的逐步完善，工业大数据迎来重大发展机遇。工业行业对大数据技术的认知和实践在几年间快速积累，技术基础设施和能力不断完善，工业大数据的关注焦点从建设工业大数据平台逐步转向数据应用解决方案；大数据在工业行业的应用场景从最初的生产监控到降本增效，逐步向支撑服务化转型探索。

2020 年初，工业和信息化部印发《工业数据分级分类指南（试行）》（工信部信发〔2020〕6 号），以指导企业提升工业数据管理能力，促进工业数据的使用、流动与共享，释放数据潜在价值，赋能制造业高质量发展。4 月印发《工业和信息化部关于工业大数据发展的指导意见》（工信部信发〔2020〕67 号）从加快数据汇聚、推动数据共享、深化数据应用、完善数据治理、强化数据安全、促进产业发展六个方面全盘布局，系统推进，针对我国工业大数据现阶段的发展特点、主要问题和亟待取得突破的重点领域，设置重点任务，精准施策，为工业大数据落地提供良好的政策环境，务实有序推动发展。

随着网络化协同、个性化定制、供应链金融等新业态新模式的快速发展，“十四五”期间，我国工业大数据将呈现如下的发展趋势:

一是积累和拓展更多的应用场景。随着工业信息化基础的增强，越来越多的工业企业开始具备数据的积累和掌控能力，比如能源电力、轨道交通、装备制造、航空航天等等。工业大数据分析大量应用到生产环节中，如设备健康管理、生产管理优化、生产监控分析、全流程系统性优化、质量管理等，通过大数据赋能企业的生产，保障生产安全、降低生产成本、提升生产效率和产品质量。

二是融合更多相关技术。随着工业互联网基础设施的建设，5G、物联网、边缘计算、区块链等技术将逐步融入工业领域的升级转型，更高质量地支撑工业大数据采集、汇聚、流通、分析、应用的价值闭环，使工业大数据的数据获取量更大、数据存储管理更便捷、数据分析产出更智能，以此助力更高效的产品工艺研发，带动服务模式创新，实现最大化的数据价值变现。

（四）互联网大数据助推商业模式创新拓展

互联网行业拥有的数据优势得天独厚。一方面，随着移动信息技术的不断进步，越来越多、种类各异的互联网应用迅速落地，使得互联网行业自身便可产生大规模、多维度、高价值的数据资源；另一方面，互联网为传输数据而生，在“互联网+”的新经济形态下，各行业产生的数据资源大都要借助互联网技术进行流通、共享与交互，互联网因此汇聚了大规模的数据，并极大促进了数据要素的价值传导。

作为大数据应用落地成型最早的行业，互联网企业深耕于如何将大数据资源转化为商业价值，在大数据的助推下进行商业模式的创新

及业务的延伸，提升用户体验，进行精细化运营，提高网络营销效率。以精准营销为典型代表的互联网大数据应用正有力推动着企业升级思维，创新模式，以数据驱动重构商业形态。

专栏 3：互联网大数据重构商业营销模式

互联网大数据很大程度地改变了传统意义的营销手段。已往的营销主要依赖品牌推广，根据群体解析；而大数据分析挖掘通过用户数据分析，市场趋势解析、触达场景解析、营销推广产品评析，洞悉营销推广对象的诉求点，利用智能推荐技术，实现了真实意义上的人性化精准营销。同时，互联网大数据还实现了线下门店和线上营销渠道的结合，让传统意义的营销手段直接进入到多屏时代。

例如，某电商平台通过客户的网络浏览记录和购买记录等对客户的收入、家庭结构、购买偏好等进行消费行为分析与预测。从消费者进入网站开始，平台在列表页、单品页、购物车页等 4 个页面部署了 5 种应用不同算法的推荐栏为其推荐感兴趣的商品，从而提高商品曝光率，促进交叉和向上销售。引入大数据进行精准营销后，平台下定订单转化率增长了 66.7%，下定商品转化率增长了 18%，总销量增长了 46%。

2020 年国务院政府工作报告中提出要“全面推进‘互联网+’”，打造数字经济新优势，互联网与大数据融合应用的能效将得到进一步发挥。“十四五”期间，我国互联网大数据呈现出如下的发展趋势：

一是互联网大数据的规模将迅速扩大，维度更加丰富。“十四五”期间将进入万物互联时代，消费互联网向产业互联网转型，物联网蓬

勃发展，对人的特征和行为数据的刻画维度将更加丰富，业务数据规模同步增大，加之各类传感器的普及使得物的数据被大量采集，互联网数据量级将迈上新的台阶。

二是技术的不断发展将推动互联网数据加速流通。隐私计算、区块链等数据流通技术的快速发展增强了数据流通过程的安全性，为数据要素的可信流通提供了有力支撑。“互联网+”的内核是数据流动，作为信息技术能力和数据资源的富矿，互联网大数据应用将进一步消除壁垒，在行业内部、跨行业间加速数据流通与融合应用。

（五）金融大数据应用成为行业核心竞争力

在全球数字化转型的热潮之中，金融行业一马当先。金融机构具有庞大的客户群体，企业级数据仓库存储了覆盖客户、账户、产品、交易等大量的结构化数据，以及海量的语音、图像、视频等非结构化数据。这些数据背后都蕴藏了诸如客户偏好、社会关系、消费习惯等丰富全面的信息资源，成为金融行业数据应用的重要基础。

随着金融业务与大数据技术的深度融合，数据价值不断被发现，有效促进了业务效率的提升、金融风险的防范、金融机构商业模式的创新以及金融科技模式下的市场监管。目前，金融大数据已在交易欺诈识别、精准营销、黑产防范、信贷风险评估、供应链金融、股市行情预测等多领域的具体业务中得到广泛应用。大数据的应用分析能力，正在成为金融机构未来发展的核心竞争要素。

专栏 4：金融大数据成为智能风控的重要基石

智能风控是金融领域大数据应用的最主要手段，通过建立大数据风险控制模型，分析客户所需承担的信用风险，实时计算统计结果，实现对客户的精准信用评分和征信结果。

随着与人工智能等技术的结合，大数据风控进一步实现了业务流程的智能化，大大提高了金融机构的效率和风控能力。根据《中国智能风控市场调研报告》，随着信贷、保险等领域业务需求的日益强烈，我国智能风控行业市场规模逐年扩张，从 2015 年的 19.1 亿元增至 2019 年的 51.2 亿元，预计 2020 年规模达 78 亿元。

毋庸置疑，金融大数据拥有广阔的发展前景，然而金融大数据应用也面临着数据质量不高、合规和安全风险严峻等一系列制约因素。“十四五”期间，我国金融大数据将呈现如下的发展趋势：

一是跨机构数据融合需求更加强烈。长期以来，金融机构在自身经营发展的过程中积累了大量的业务数据，但是内部数据往往来源单一、片面，无法支撑多样、全面、深入的数据分析需求。为了能够全面洞察客户需求、预测经营风险、支撑经营决策，引入外部数据实现跨机构间的数据融合变得尤为重要。

二是数据安全与合规管理更加重要。中国信通院发布的《2018-2019 年度金融科技安全分析报告》显示，针对客户资料与企业重要业务数据的安全事件是金融机构发生频率最高的风险事件，占比 44%。随着监管部门对数据安全和个人隐私保护日益重视，在强监管态势下，金融行业如何保障数据安全及合规使用，已成为当前亟需解决的问题。

五、数据治理

关于数据治理的概念界定，一般有广义和狭义之分。狭义的数据治理主要是指对数据进行治理的技术与活动，是组织内部对数据的处置与应用进行规范化的过程；而广义的数据治理则是通过多样化治理手段激活与释放数据要素价值的一套行为体系，是发展数字经济的关键所在。

从广义的角度，我们认为**数据治理是企业、政府、社会、市场等多参与主体，通过技术、制度、人员、法律等多种方式，实现提升数据质量与应用价值、促进数据资源整合与流通共享、保障数据安全等目标的一整套行为体系**。在数据治理的实施过程中，组织内部的数据管理，组织间社会化的数据流通和覆盖数据全生命周期的安全保障是三个关键议题。本章将针对以上三点的现状与热点进行简要介绍。

（一）组织内部的数据管理能力逐步提升

在企业和政府的数字化转型过程中，数据是基础性和战略性资源。只有夯实数据管理之基，才能提升数据资源质量，支撑上层的数据流转与应用，充分发挥数据资源的价值。

1. 企业积极实践数据资产管理

不同行业的数据资产管理实践模式有所差异。经过多年发展，企业数据资产管理的理论基础已逐步成熟，形成了以国际数据管理协会（DAMA）的数据管理模型、数据治理研究所（DGI）的数据治理框架等为代表的理论框架，我国也于 2018 年发布了国家标准 GB/T 36073-2018《数据管理能力成熟度评估模型》，简称 DCMM。但在

各行业的具体实践中，理论的共性逐渐被行业的个性所替代。例如，**金融行业普遍“管理制度先行”，**针对性的建立数据质量部门、数据标准部门、数据开发部门、数据分析部门等相关管理部门，数据资产管理活动侧重于监管数据治理、信息系统、数据安全、应急预案。**互联网企业通常 “实践探索先行”，**将数据模型、数据仓库、数据分析作为核心应用，随着网络数据安全保护能力专项行动的开展和个人信息保护的加强，数据安全也逐渐成为互联网企业的关键数据资产管理活动。

不同行业的数据资产管理综合能力差距明显。结合 DCMM 对于数据资产管理综合能力的等级划分（如表 2 所示），与在 DCMM 贯标与评估工作中对各行业实践现状的观察与总结可以发现，**金融、电信、互联网等行业的数据资产管理综合能力多处于稳健级和量化管理级，其它行业多数仍处于初始级和受管理级。**

表 2 数据资产管理能力划分[9]

	数据战略	数据治理	数据架构	数据应用	数据安全	数据质量	数据标准	数据生存周期
初始级	仅在项目范围建立了数据战略	项目层面实现了数据治理	应用系统层面建立了数据架构	项目层面开展数据应用	项目层面开展数据安全管理	项目层面制定了数据质量提升方案	定义了项目范围的数据标准	项目层面开展数据生存周期管理
受管理级	数据战略和业务战略关联并实施	部门层面明确了数据职责和数据制度	针对数据管理具体问题构建相对完善的数据架构	部门层面多以线下方式进行数据应用	部门层面建立了数据安全标准和管理策略	部门层面识别了关键数据质量需求并逐步提升	定义了部门范围的数据标准并逐步落地标准	部门层面明确数据需求、数据设计开发、数据运维和数据退役规范

[9] 此处数据治理为狭义概念。

稳健级	建立反映整个组织发展的数据战略优化路线图并实施	组织层面建立了完善的数据组织、数据制度体系	组织层面建立了数据模型、数据分布、数据集成与共享、元数据及其管理规范	组织层面以提升数据价值为驱动力，充分利用各类技术平台全面进行数据分析、开放共享和数据服务	建立组织级数据安全团队和管理制度，充分识别和满足数据安全监管要求	确立了组织级数据质量关键问题和提升目标，制定了数据质量改善路线并实施	创建了组织级的业务术语、参考数据、主数据、数据元和指标数据及其管理制度，数据标准全面落地	数据生存周期管理覆盖了组织范围数据以及数据采集、存储、加工、退役全流程
量化管理级	基于数据战略量化指标持续跟踪优化	建立数据组织和数据制度考核指	建立了评价数据架构的量化指标体系	建立数据应用的评价指标以衡量数据价值	根据内外部环境变化，不断调整数据安全策略	形成数据质量提升方案闭环，从源头改善质量	对数据标准体系和落地效果不断改进	数据生存周期管理有效支撑数据战略和数据应用
优化级	数据战略有效提升企业竞争力并成为行业标杆	数据组织和制度有力支撑数据管理工作	参与数据架构国际和国家标准制定，持续对外输出优秀经验	数据应用为企业和社会创造价值	参与数据安全国际和国家标准制定	数据质量管理实践具有先进性	参与国际、国家、地方和行业相关数据标准制定	

来源：GB/T 36073-2018《数据管理能力成熟度评估模型》，中国信息通信研究院

金融、电信、互联网行业的数据资产管理能力优势集中于数据战略、数据治理、数据架构和数据生存周期方面。**在金融行业**，已有 67%的金融机构建立了组织级的数据治理架构并明确了相关管理层职责，85%的机构表示已经将数据治理纳入到组织年度战略规划中，67%的机构建立了全行统一的大数据分析平台[10]，并开展数据应用、数据服

[10] 数据来源：《2019 年度中国金融行业数据治理调研报告》

务相关的培训宣贯。**在电信行业，**以中国联通为例，在 2014 年就开始了组织层面的数据资产管理综合能力建设，实现了 31 个省公司指标数据和明细数据的一致，建立了包括信息化管理委员会、数据治理办公室、业务数据管理单位和数据生产单位在内的组织架构，并于 2019 年启动企业级数据中台建设，实现了 BOM 全域物理集中和逻辑集中。**在互联网行业，**以滴滴为例，在 2017 年成立了数据治理部、数据架构部、数据平台部和数据科学部，利用大数据平台对数据全生存周期的管控，并通过将数据科学家分散至各业务部门的方式，加深数据需求与业务需求的融合。此外，中国互联网协会于 2020 年 7 月成立了数据治理工作委员会，通过搭建公共平台、制定共同标准等推动互联网行业数据治理能力的提升。

其它行业数据资产管理能力不足的主要原因包括信息化基础薄弱、数据管理投入人员和专业水平不足、数据资产管理驱动力受限等。**以工业为例，在信息化基础方面，**相较于头部行业成熟的大数据平台和正在建立的数据基础设施，工业企业依然停留在 ERP、CRM、SCM 的管理软件阶段，使得企业级数据采集、存储和分析的成本较高，目前 41%的工业企业仍然使用手工或文档方式进行数据管理[11]；**在数据管理投入人员和专业水平方面，**将近一半的工业企业在数据管理环节投入的人数为 5 人以下，未建立专业的数据管理团队；**在数据资产管理驱动力方面，**数据多用于监控生产运营和设备故障，数据应用场景狭窄，缺少数据资产带动业务发展的强驱动力。

2. 政府数据管理水平不断提升

[11] 数据来源：《工业企业数据资产管理现状调查报告》（2018）

作为“数字中国”的有机组成部分，建设“数字政府”成为近年来各地积极推进的重要工作，在夯实数据基础之上，通过“用数据说话，靠数据决策，依数据行动”，增强政府决策的科学性、预见性和精准性。自 2014 年部分省市陆续成立大数据局开始，截至目前，全国共有 20 个省级政府和 80 个以上的副省级和地级市政府成立了专门的职能机构，对政府内部的数据资源进行统一协调管理。在此基础上，各地政府采取了多项举措进一步加强数据管理能力。

一方面，多地颁发的政府数据管理办法明确了数据权责清单。各级政府纷纷建立数据统筹集约管理机制，建立覆盖数据资源全生命周期的管理制度体系和标准体系。一般的，由当地政府统一领导，各地各部门按照省统一的标准规范和要求，组织开展数据治理专项工作，由大数据主管机构负责指导、监督、管理和协调工作，并对各级行政机关和企事业单位的数据资源管理情况进行考核，各级行政机关和企事业单位依据“三定权责”进行数据共享开放，并开展数据管理。

以广西为例，自 2019 年 11 月以来，自治区大数据发展局围绕广西政务数据资源管理与应用改革，加快完成政务数据“聚通用”工作。截至目前，共发布了《自治区厅局数据治理试点协同推进工作方案》、《厅局中台建设工作指引》、《数据治理实施方案》等 11 项标准规范；相关试点厅局开展的数据治理工作共梳理了 479 项权责、740 个事项、1908 条资源目录，接入 29 个系统、归集 10 亿条数据、制定 669 条数据标准，全区已基本形成一云承载初步实现、一网通达逐级贯通、一池共享雏形初显、一事通办已见成效和一体安全有序实施的“五个一”政务数据治理新模式。

另一方面，各地加强了对数据质量、数据标准、数据目录、元数

据、数据采集、数据审计和数据安全等方面的要求。包括确保数据真实准确和完整及时、按照相关技术标准对数据进行结构化处理（据统计，各地开放数据集中满足可机读格式标准的比率达到 82%[12]）、数据目录更新维护、明确元数据信息、采用"一数一源"采集原则、建立日志记录以确保数据使用过程可追溯、数据分类分级等。同时，各地政府也有序推动各类数据建库工作，深化基础数据库共建共用，促进政务数据共享开放，充分挖掘分析公共数据潜在价值。

以海南为例，2019 年，海南省政府基于当时政务大数据平台数据采集全面性不足、数据实时性不够、数据质量有待提升、数据服务出口不统一、数据完整性较差、缺少实时数据服务等现状，提出建设基础数据质量提升工程。截至目前，海南省政务大数据公共服务平台已完成元数据管理系统、数据质量管理系统、数据标准管理系统、数据分析系统、共享体系等系统的完善升级，构建了数据指标管理系统与标签管理系统，并实现了 317 个政务系统，43037 张表，77 万个信息项，48 亿条记录数的归集。

（二） 组织间的数据共享与流通加速推进

作为重要的生产要素，在提升组织内部数据管理能力的基础之上，政府与企业之间、企业与企业之间自由有序的数据流通与共享亦是数据要素价值充分释放的关键。

1. 各地政府数据开放共享效果显著

相对企业数据来说，政府的数据资源规模更大、种类更多，且事

[12] 数据来源：《2020 联合国电子政务调查报告》

关百姓生活的方方面面，数据的潜在价值巨大，政府数据的开放将为推进数字经济高质量发展提供巨大能量。

一方面，各地政府数据开放共享的制度体系逐步完善。根据中国信通院统计，截至目前，除黑龙江以外，全国共有 30 个省份出台了 56 份政府数据开放的相关政策文件（如表 3 所示）。**在确定数据共享开放内容方面，**各地坚持需求导向、目标导向，征求行业协会、相关企业、社会公众和行业主管部门的意见建议，形成数据开放目录清单，并动态调整。**在共享开放数据范围方面，**由政府数据扩展至企事业单位所涉及的公共数据资源，据统计，全国开放数据集总量从 2017 年到 2019 年增加了 6 倍[13]。**在共享开放数据成果形式方面，**提供包括服务应用、数据可视化、研究成果、创新方案等数据共享开放形式。

表 3 各省（区、市）政务数据开放相关政策文件

地区	政策文件	发布时间
贵州	《贵州省政府数据共享开放条例》	2020.09.25
	《贵州省政务信息系统整合共享工作方案》	2017.10.13
	《贵州省政务数据资源管理暂行办法》	2016.11.01
重庆	《重庆市公共数据开放管理暂行办法》	2020.09.11
	《重庆市政务数据资源管理暂行办法》	2019.07.31
	《重庆市政务信息资源共享开放管理办法》	2018.05.05
	《重庆市政务信息系统整合共享工作方案》	2017.10.09
天津	《天津市公共数据资源开放管理暂行办法》	2020.07.21
浙江	《浙江省公共数据开放与安全管理暂行办法》	2020.06.12
	《浙江省公共数据和电子政务管理办法》	2017.05.01
广西	《加快推进广西政务信息系统整合共享实施方案》	2020.03.23
	《广西政务数据资源管理与应用改革实施方案》	2019.11.02
	《广西政务信息资源共享管理暂行办法》	2016.10.21
山东	《山东省电子政务和政务数据管理办法》	2019.12.25
	《山东省政务信息系统整合共享实施方案》	2017.10.22
	《山东省政务信息资源共享管理办法》	2015.03.01

[13] 数据来源：《2020 联合国电子政务调查报告》

辽宁	《辽宁省政务数据资源共享管理办法》 《辽宁省政务信息系统整合共享实施方案》	2019.11.26 2017.08.16
上海	《上海市公共数据开放暂行办法》 《上海市公共数据和一网通办管理办法》 《上海市政务数据资源共享管理办法》	2019.08.29 2018.09.30 2016.02.29
吉林	《吉林省公共数据和一网通办管理办法（试行）》 《吉林省加快推进政务信息系统整合共享工作方案》	2019.01.04 2017.11.22
广东	《广东省政务数据资源共享管理办法（试行）》 《广东省政务信息系统整合共享工作方案》	2018.11.29 2017.10.30
湖北	《湖北省政务信息资源共享管理办法》	2018.09.26
宁夏	《宁夏回族自治区政务数据资源共享管理办法》 《宁夏回族自治区政务信息系统整合共享分工方案》	2018.09.04 2017.10.09
内蒙古	《内蒙古自治区政务信息资源共享管理暂行办法》	2018.06.15
海南	《海南省公共信息资源管理办法》 《海南省政务信息整合共享专项行动实施方案》	2018.05.25 2017.09.30
河南	《河南省政务信息资源共享管理暂行办法》 《河南省政务信息系统整合共享实施方案》	2018.01.08 2017.09.29
北京	《北京市政务信息资源管理办法（试行）》	2017.12.27
西藏	《西藏自治区政务信息系统整合共享实施方案》	2017.12.22
云南	《云南省政务信息资源共享管理实施细则》	2017.12.22
湖南	《湖南省政务信息系统整合共享实施方案》 《湖南省政务信息资源共享管理办法（试行）》	2017.12.19 2017.11.22
江苏	《江苏省政务信息资源共享管理暂行办法》 《江苏省政务信息系统整合共享工作实施方案》	2017.10.26 2017.09.07
福建	《福建省政务信息系统整合共享实施方案》 《福建省政务数据管理办法》 《福建省政务信息共享管理办法》	2017.10.25 2016.10.15 2010.11.12
山西	《山西省政务信息系统整合共享工作方案》	2017.10.12
甘肃	《甘肃省政务信息系统整合共享实施方案》	2017.09.30
河北	《河北省政务信息系统整合共享实施方案》 《河北省政务信息资源共享管理规定》	2017.09.22 2015.11.13
陕西	《陕西省政务信息系统整合共享实施方案》 《陕西省政务信息资源共享管理办法》	2017.09.13 2017.08.16
江西	《江西省政务信息系统整合共享实施方案》 《江西省政务信息资源共享管理实施细则》	2017.09.13 2017.01.16

安徽	《安徽省政务信息系统整合共享实施方案》 《安徽省政务信息资源共享管理暂行办法》	2017.08.23 2017.02.21
青海	《青海省政务信息系统整合共享工作方案》	2017.08.01
新疆	《关于推进新疆维吾尔自治区政务信息资源共享管理工作的实施意见》	2017.05.20
四川	《四川省政务信息资源共享管理实施细则（暂行）》	2017.03.05

来源：根据公开资料整理，中国信息通信研究院

另一方面，各地政府数据开放共享的落地实施进展加快。截至 2020 年 4 月底，全国已有 130 个省级、副省级和地级政府上线了数据开放平台[14]，具备包括数据检索、数据申请、数据获取和数据探索等基本能力，形成了省市分级维护数据资源、协同共享数据资源体系的局面。以深圳市政府数据开放平台为例，从 2016 年 11 月上线以来，截至 2020 年 7 月，平台访问次数近 2478 万次，数据总量近 2.7 亿条，内容涉及教育科技、交通运输等 14 个领域。

同时，多地举办数据创新应用活动，推动政府数据开发利用，激发政府数据活力。深圳、贵阳、上海、杭州、成都等地连续举办了开放数据应用创新大赛，在进一步推动政府部门开放数据的同时，也提升了政府数据资源利用效率，激发社会创新活力。以深圳市 2020 年开放数据应用创新大赛为例，开放了 2 亿 7 千条脱敏后的真实数据，涉及疫情防控、环境保护、社会治理、医疗健康等多个领域，诞生了如"疫情防控下空间智能支持居民生活保障"、"城市地摊经济解决方案"等数据利用成果。

2. 数据交易市场建设正在努力探索

要实现政府、企业等组织外部社会化的、自由有序的数据交易与

[14] 复旦 DMG，2020 中国政府数据开放报告，https://mp.weixin.qq.com/s/tKt8wvM0X4p0MFzpEDlBxg

流通需要建立规范有序的交易市场来提供健康发展的土壤。

2014年：贵阳大数据交易所；北京大数据交易服务平台
2015年：淮南大数据交易平台；重庆大数据交易平台；河北大数据交易中心；长江大数据交易中心；华东江苏大数据交易平台；武汉东湖大数据交易中心；西咸新区大数据交易所
2016年：中原大数据交易平台；钱塘大数据交易中心；华中大数据交易平台；哈尔滨数据交易中心；浙江大数据交易中心；上海大数据交易中心
2017年
2018年
2019年：山东数据交易公司
2020年：粤港澳大湾区数据平台；湖北大数据交易中心；北方大数据交易中心；湖南大数据交易中心；北京国际大数据交易所；北部湾大数据交易中心

来源：中国信息通信研究院

图 14　国内大数据交易市场建设历程

我国的数据交易产业起步于 2014 年，随着我国大数据元年的到来一同开启，各地政府积极支持。2014-2016 年间，国内大数据交易所呈井喷态势，各地开花，不到三年的时间里先后成立了 13 家大数据交易所（中心、平台）。但几年时间过去，各家交易所的运营情况大多不尽如人意，数据交易的成交量远低于预期设想，甚至很多已经陷入搁置、停运状态，数据交易产业仍处在小规模探索阶段。

究其原因，主要在于数据交易所的定位和模式未明、数据交易配套的法律痼疾未祛。一方面，各交易所建设时的定位相似、功能重复，在缺少核心竞争优势的同时，服务模式、定价标准等交易规则体系参差混乱，难以培养数据供需双方对交易所的平台依赖，只能沦为小规模数据交易的撮合者。另一方面，数据权属的界定仍处于灰色地带，在相关立法尚未健全的当下，行业内的实践中并未能形成具有共识性或参考性的权属分割规则，产权争议、无法监管的风险令供需双方望而却步。除此之外，频发的数据安全和个人隐私泄露事件加剧了社会对数据交易的不信任感，出于对国家安全、个人信息和商业秘密的保护，主体参与数据交易的主动性、积极性因此降低，成为数据交易所

发展的又一大障碍。

随着中央提出加快培育数据要素市场的愿景，以及新的市场环境和技术条件，大数据交易市场又出现了新的生机，我国国内的数据交易产业重新起航。

自 2019 年年底以来，各地重新布局数据交易产业的脚步加快。在各地 2020 年政府工作报告中，湖北省提出将筹建湖北大数据交易集团，天津滨海新区提到要加快建设北方大数据交易中心。2019 年 12 月 11 日，**山东数据交易公司**在济南揭牌成立，并联合上海、江苏、安徽的数据交易机构共同发起成立华东数据联盟，孵化跨省市的数据交易流通生态。2020 年 7 月 14 日，**湖南大数据交易中心**正式开工建设，预期年内完成基本建设。8 月 12 日，**北部湾大数据交易中心**在南宁揭牌，作为国际化数据资源交易服务机构，面向中国与东盟区域提供全链条、一站式的数据服务。

北京探索建设国际大数据交易所。2020 年 9 月 5 日，北京市委宣布将建设国际大数据交易所。9 月 7 日，《北京国际大数据交易所设立工作实施方案》正式发布，其建设目标定位于国内领先的大数据交易基础设施和国际重要的大数据跨境交易枢纽。

北京国际大数据交易所明确了五大功能定位: 一是权威的数据信息登记平台，通过充分的信息披露明晰数据的取得方式及权利范围；二是受到市场广泛认可的数据交易平台，健全报价、询价、竞价和定价机制，对数据产品的所有权交易、使用权交易、收益权交易和跨境交易进行划分并提供多种交易模式；三是覆盖全链条的数据运营管理服务平台，提供数据清洗、法律咨询、价格评估、分析、评议、尽职调查等全链条的服务；四是以数据为核心的金融创新的服务平台，探

索开展基于数据资产的质押融资、保险、担保、证券化等金融创新服务；五是新技术驱动的数据金融科技平台，深入挖掘多方安全计算、区块链等技术在数据安全、数据应用等方面的作用并充分运用。

中央支持深圳建设粤港澳大湾区数据平台。2020 年 10 月 11 日，中共中央办公厅、国务院办公厅发布《深圳建设中国特色社会主义先行示范区综合改革试点实施方案（2020-2025 年）》，支持深圳在更高起点、更高层次、更高目标上推进改革开放。其中在数据要素方面，提出了“率先完善数据产权制度，探索数据产权保护和利用新机制，建立数据隐私保护制度。试点推进政府数据开放共享。支持建设粤港澳大湾区数据平台，研究论证设立数据交易市场或依托现有交易场所开展数据交易。开展数据生产要素统计核算试点”的要求，为深圳乃至整个粤港澳大湾区数据要素市场化配置提供了整体思路。

表 4　新设大数据交易所（中心）的基本情况

	建设进程	功能定位	服务模式	覆盖的数据资源
山东数据交易公司	已揭牌成立	省级综合性数据服务平台	提供数据交易平台、数据产品开发、数据应用、公共数据资源开放及其他类型等五大类服务	以公共数据资源为基础，吸引商业数据资源；依托华东数据联盟，吸引跨省数据资源
湖南大数据交易中心	预期年内完成场地的基本建设	具有全国、全球影响的数据集聚、流通、应用的数据资源交易场所	采取“淘宝”的运营模式，为有交易需求的数据资源提供数据存储、定价、交易、监管服务	作为国家地理空间信息中心的南部分中心，汇聚湖南、福建、广东、海南等南方九省的地理信息空间数据资源，并以此为基础，整合政务数据、通信运营商数据、互联网企业数据、金融数据等；
北部湾大数据交易中心	已揭牌成立	以“政府指导，自主经营，市场化运作”为原则的国际化数	以交易佣金、授权使用费、资源使用费、定制产品开发费、数	面向国内和东盟地区各国汇聚、处理、使用和交易各类数据产品

		据资源交易服务机构和数据服务全生态交易平台	据深度加工服务费等为盈利模式	
北京国际大数据交易所	启动建设（由具有优质数据资源的市属国企对现有交易所进行重组、更名）	数据信息登记平台；数据交易平台；数据运营管理服务平台；数据资产金融创新服务平台；数据金融科技平台	基于五大功能定位提供数据信息登记服务、数据产品交易服务、数据运营管理服务、数据资产金融服务和数据资产金融科技服务	北京市政府部门(含具有公共事务职能的组织)将数据目录中的公共数据通过无条件开放和授权开放形式有序汇聚；同时，驱动商业数据聚集

来源：根据公开资料整理，中国信息通信研究院

（三） 数据安全治理成为不可忽视的焦点

作为生产要素，数据的需求与应用日益广泛，数据要素价值的释放路径更加多元，但无论是组织内部的数据应用还是组织间的数据流通，数据面临的安全风险也随着其价值的逐步凸显而更加突出。一方面，数据应用的复杂性和数据分析挖掘的多样性增加了数据权属管理和抵御安全攻击的难度；另一方面，越来越多的跨组织间数据流通进一步加速了数据被盗用、误用、滥用的安全风险。近年来，数据安全事件的层出不穷使得数据安全治理也成为各界无法忽视的焦点问题。

1. 数据安全标准制度体系逐步构建

数据安全管理的相关标准制度不断完善。2019 年 8 月 30 日，《信息安全技术 数据安全能力成熟度模型》（GB/T 37988-2019）简称 DSMM（Data Security Maturity Model）正式成为国标对外发布，并已于 2020 年 3 月起正式实施。DSMM 从数据采集安全、数据传输安全、数据存储安全、数据处理安全、数据交换安全、数据销毁安全六个维度提出了覆盖全生命周期的数据安全能力要求，为各类组织开展数据

安全治理提供指引。

除国家标准外，工信部于 2020 年 3 月 4 日印发《工业数据分类分级指南（试行）》，指导工业领域产品和服务全生命周期产生和应用的数据的分类分级工作。更进一步地，于 4 月 10 日发布《网络数据安全标准体系建设指南》（征求意见稿），提出了“到 2021 年，初步建立网络数据安全标准体系”，“到 2023 年，健全完善网络数据安全标准体系”的标准制度建设目标。中国人民银行于 2020 年 9 月 23 日发布《金融数据安全 数据安全分级指南》（JR/T 0197—2020）金融行业标准，明确了数据安全定级的要素、规则和定级过程，并给出了金融业机构典型数据定级规则供实践参考。

行业级数据安全合规的监管行动正在积极开展。2019 年 7 月 1 日，工信部印发《电信和互联网行业提升网络数据安全保护能力专项行动方案》，开展为期一年的专项行动，集中开展数据安全合规性评估、专项治理和监督检查，推动建立行业级的网络数据安全保障体系，促进电信和互联网行业提升数据安全管理提升，有力推动了行业数据安全治理体系的构建。2020 年 5 月 14 日，工信部发布了《2020 年电信和互联网企业网络数据安全合规性评估要点》，明确了行业数据分类分级、安全评估、安全认证、预警处置等关键制度规范和要求。

2. 企业数据安全治理实践逐步深入

频发的数据安全事件为企业造成了巨大损失，在标准制度提供指引和政府监管力度不断增强的背景之下，各行业企业对于数据安全治理的自主意识不断增强，企业级的数据安全治理实践不断深入。

一方面，重视数据安全制度建设成为各类企业共识。实现数据安

全治理，需要制度流程、技术手段与组织人员三位一体、高度融合。其中，建立完善的数据安全管理制度是重要基础。在实践过程中，各行业企业纷纷建立起了覆盖“事前预防-始终监控-事后处置”全流程的制度规范体系。

以小米集团为例，在小米的安全保障体系中，建立了事前安全防范、事中安全管控、事后稽核审计的一整套完整安全管理制度。其中，事前安全防范包括制定数据安全管理细则与完善审批流程、大数据资产管理、数据分类分级制度与合作方调研审查；事中安全管控包括数据权限管理、数据共享管理、合作方数据安全管理与个人信息保护；事后稽核审计包括数据安全审计、数据安全预警与应急处置。可以看到，在这样的数据安全制度体系中，不仅针对企业内部数据的管理与应用进行管控，对于企业外部数据的流通共享也建立了针对性的管理制度，可以覆盖到数据价值流转的各个环节。

另一方面，数据安全技术应用是保障企业数据安全的重要支撑。在完善制度体系的基础之上，建立完整可靠的数据安全技术体系更是抵御内外部数据安全风险的关键。在传统信息安全涉及的网络安全、主机安全和应用安全的基础之上，覆盖数据采集、存储、挖掘到销毁等全生命周期的技术安全保障更为重要。

以联通为例，联通大数据在企业数据安全治理实践中建立了完整的技术保障体系。在数据采集环节进行数据分类分级、身份认证、数据加密；在数据存储环节进行数据脱敏、细粒度权限与访问控制、多副本多节点备份；在数据挖掘环节对数据使用行为进行监测；在数据发布环节进行内容审计、数据溯源与合规管控；在销毁环节进行定期销毁。此外，联通通过开发数据资产地图、数据安全网关、全息审计

平台等配套产品保障了企业内部数据安全治理实践的技术应用。

除此之外，多方安全计算、可信执行环境等“数据可用不可见”的隐私计算技术也在企业间数据流通环节中发挥了重要作用，帮助企业数据安全治理构建了有效的数据闭环。

六、大数据法制

大数据产业在发挥资源禀赋效应的同时，也催生出诸多隐患，如侵犯个人信息和隐私，泄露国家秘密，数据被截获、篡改和伪造，数据权属不明，数据垄断以及不正当竞争等[15]。近年来我国数据立法进程不断加快。2020 年 7 月，《中华人民共和国数据安全法（草案）》面向公众公开征求意见。作为数据安全领域的上位法，《数据安全法（草案）》在数据分级分类、监测预警和应急处置等数据安全相关的各项管理制度和数据安全责任体系构建等方面提出要求。但从目前的条款设置来看，可实际操作落地的具体化规定仍然不足，在如何划分重要数据的权属边界、数据交易如何进行等问题还有待进一步落实。

推进大数据产业的创新发展，必须加强领域内的法律建设，构建公平、自由、有序的市场竞争环境。本章将重点对近年来我国数据立法的主要着眼点进行梳理。

（一）强化个人权益，个人信息保护立法加快

数据立法缺失反映最为突出的问题是非法获取、分享和交易导致的个人信息泄露与滥用。骚扰电话、短信、邮件泛滥，个人财产损失和名誉损害事件屡见不鲜。下表列举了我国近年来涉及个人信息保护的主要规范。

表 5 个人信息保护相关的主要规范

生效时间	文件名称	发布部门
2012	《关于加强网络信息保护的决定》	全国人大常委会
2013	《消费者权益保护法》	全国人大常委会

[15] 詹馥静,王先林：《反垄断视角的大数据问题初探》，载《价格理论与实践》2018 年第九期，第 37 页。

2013	《电信和互联网用户个人信息保护规定》	工业和信息化部
2014	《网络交易管理办法》	国家工商行政管理总局
2015	《中华人民共和国刑法修正案(七)》	全国人大常委会
2017	《中华人民共和国网络安全法》	全国人大常委会
2017	《关于办理侵犯公民个人信息刑事案件适用法律若干问题的解释》	最高人民法院 最高人民检察院
2019	《App 违法违规收集使用个人信息行为认定方法》	国家互联网信息办公室、工业和信息化部办公厅、公安部办公厅、国家市场监督管理总局办公厅
2019	《儿童个人信息网络保护规定》	国家互联网信息办公室
2019	《国务院关于在线政务服务的若干规定》	国务院
2019	《中华人民共和国电子商务法》	全国人大常委会
2021	《民法典》	全国人民代表大会

来源：根据公开资料整理，中国信息通信研究院

2020 年 10 月，《中华人民共和国个人信息保护法(草案)》公布，草案确立了个人信息处理应遵循的原则，即强调处理个人信息应当采用合法、正当的方式，具有明确、合理的目的，处理信息应当遵循公开、透明的原则。草案在个人信息处理规则的章节明确了个人信息处理的主要合法性基础，包括：（一）取得个人的同意;（二）为订立或者履行个人作为一方当事人的合同所必需;（三）为履行法定职责或者法定义务所必需;（四）为应对突发公共卫生事件,或者紧急情况下为保护自然人的生命健康和财产安全所必需等；这些合法情景结合了实践中数据处理的多种场景，例如：此次疫情中政府提供的"健康码"服务，其数据处理的合法性可以满足第四项要求；此外，这一章节还从共同处理、委托处理、合并分立、与第三方合作处理等多个角度规定了个人信息处理的具体规则。

《草案》首次将国家机关全面纳入到个人信息保护法规制范围内，是目前草稿最大的亮点。专门增加一节规定：国家机关处理个人信息

的，应当具有法定职权，并按照法定权限、程序进行。这将有助于整体社会的个人信息保护水平，并推动国际社会对中国个人信息保护制度的认可，为后续推进跨境数据流动机制提供了积极条件。

此外，《草案》还对各界普遍关注的敏感个人信息进行了特殊保护，即要求在具有特定目的和充分必要性的前提下，获得个人的单独或书面同意，另外还应当告知处理个人敏感信息的必要性及对个人的影响。

（二）坚持多边合作，数据跨境流动立法加强

在数字经济发展的今天，数据只有实现在更大范围内的流动共享，才能更好地发挥对经济增长、社会发展、全球化进程的支撑推动作用。随着经济全球化发展进程的加快，数据的跨境流动需求日益增强，必须在法规制度、责任体系、安全风险防范等方面上做好保障。

当前，世界多国通过国内立法或国际协定等方式快速推进数据跨境流动规则制定。例如，2019 年 9 月，日本与美国签署的贸易协定提到，“确保各领域数据无障碍跨境传输”以及“禁止对金融业在内的机构提出数据本地化要求”，希望制定旨在促进数据自由流动的规则。特别是在新冠疫情背景下，各国在数据跨境流动领域频繁发力[16]。2020 年 3 月，基于《合法使用境外数据明确法》，澳大利亚联邦政府修订《电信（拦截和接入）法案》，允许协议国在出于执法目的时，互相跨境访问通信数据。2020 年 6 月，英国宣布脱欧后的未来科技贸易战略，允许英国和某些亚太国家间的数据自由流动，并希望与日本等国达成比其作为欧盟成员国时期更进一步的数据协议。欧盟最高法院

[16]姜颖：《全球数据跨境流动发展态势及建议》，中国信息通信研究院政策与经济研究所。

出于对欧盟公民数据隐私安全的考虑，于 2020 年 7 月宣布废除《隐私盾》（Privacy Shield）跨大西洋数据保护协议。

近年来，我国加强了数据跨境流动领域决策部署，以《网络安全法》为基础，进一步进行了制度探索，有关情况如下表所示：

表 6　我国数据跨境流动的相关政策法规及标准情况

名称	发布时间	发布部门	相关条文
《数据安全法(草案)》	**2020.07.03**	全国人大常委会	第十条　国家积极开展数据领域国际交流与合作，参与数据安全相关国际规则和标准的制定，促进数据跨境安全、自由流动。
《海南自由贸易港建设总体方案》	**2020.06.01**	中共中央国务院	11.便利数据流动。在国家数据跨境传输安全管理制度框架下，开展数据跨境传输安全管理试点，探索形成既能便利数据流动又能保障安全的机制。
《信息安全技术　个人信息安全规范》（GB/T 35273—2020）	**2020.03.06**	中华人民共和国国家质量监督检验检疫总局 中国国家标准化管理委员会	9.8 个人信息跨境传输 在中华人民共和国境内运营中收集和产生的个人信息向境外提供的，个人信息控制者应遵循国家相关规定和相关标准的要求。
《中国(上海)自由贸易试验区临港新片区总体方案》	**2019.07.27**	国务院	（九）实施国际互联网数据跨境安全有序流动。试点开展数据跨境流动的安全评估，建立数据保护能力认证、数据流通备份审查、跨境数据流通和交易风险评估等数据安全管理机制。
《个人信息出境安全评估办法》	**2019.06.13**	国家互联网信息办公室	全文
《数据安全管理办法(征求意见稿)》	**2019.05.28**	国家互联网信息办公室	第二十八条 向境外提供个人信息按有关规定执行。 第二十九条 境内用户访问境内互联网的，其流量不得被路由到境外。
《信息安全技术 数据出境安全评估指南(征求意见稿)》	**2017.08.30**	全国信息安全标准化技术委员会	全文
《个人信息和重要数	**2017.04.11**	国家互联网信息	全文

据出境安全评估办法（征求意见稿）》		办公室	
《网络安全法》（中华人民共和国主席令第 53 号）	**2016.11.07**	全国人大常委会	第十二条 国家保护公民、法人和其他组织依法使用网络的权利，促进网络接入普及，提升网络服务水平，为社会提供安全、便利的网络服务，保障网络信息依法有序自由流动。第三十七条 关键信息基础设施的运营者在中华人民共和国境内运营中收集和产生的个人信息和重要数据应当在境内存储。因业务需要，确需向境外提供的，应当按照国家网信部门会同国务院有关部门制定的办法进行安全评估；法律、行政法规另有规定的，依照其规定。

来源：根据公开资料整理，中国信息通信研究院

在《网络安全法》出台之前，我国未对跨境数据流动作出系统性规范。《网络安全法》第三十七条原则上禁止我国境内关键基础设施运营者向境外提供在中国境内收集和产生的个人信息或重要数据，可见我国目前对于关键领域数据保护采取了倾向数据本地化存储的态度。但我国并非完全禁止跨境数据流动，《网络安全法》第三十七条对向境外提供个人信息和重要数据的活动设置了安全评估的前提性要求，把安全评估办法的制定权力赋予国家网信部门，其可会同国家有关部门对安全评估制定具体的操作制度。

我国在加强数据跨境流动法律建设的同时也在推动试点以探索合理的跨境流动方式。2020 年 6 月 1 日，中共中央、国务院印发《海南自由贸易港建设总体方案》，标志着海南自由贸易港的建设进入全面实施阶段。方案指出要建立健全数据出境安全管理制度体系、健全数据流动风险管控措施。国务院于 2019 年 8 月 6 日发布的《中国（上海）自由贸易试验区临港新片区总体方案》也指出，试点要开展数据

跨境流动的安全评估，建立数据保护能力认证、跨境数据流动和交易风险评估等数据安全管理机制。

（三）明确权利属性，数据权属立法探索初现

大数据创造了财富，但现在面临着数据权属不清楚因而收益分配不清楚的现实问题，这反映了加快数据所有权研究和立法的紧迫性和必要性。

《深圳经济特区数据条例》首次规定了数据权，并被纳入 2020 年拟新提交审议项目（以下简称《条例》）。《条例》在数据确权、组织机构、公共数据开放利用等方面进行了大胆的创新和尝试。然而，《条例》存在超权立法的可能性。根据《立法法》第八条，涉及民事基本制度和基本经济制度，以及对非国有财产的征收、征用的事项，只能通过制定法律给予规制。但《条例》第一百零一条规定，个人数据包括个人信息数据和隐私数据，直接将个人数据权属赋予了人格权属性，涉及我国民事基本制度。第四条规定，数据权包括权利人对数据的处分、收益、损害赔偿权力，给数据权赋予了财产权力的属性，涉及基本经济制度。

“数据权属”问题是一直影响数据资产化、数据交易的老大难问题，这主要是因为数据有着不同于土地、资本等传统生产要素的特点——数据不仅是生产要素，也附着了社会关系，各方主体对数据的权益都有所投射，在数据处理周期中，难以将权属归于单一的主体。

需要看到的是，“数据权属”虽没有明确定论，但并不会构成对数据开放利用的阻碍；相反，数据价值开发在根本上并不取决于传统的所有权定性，而是通过多方的市场参与，达成数据共享利用、促进价

值生成的市场共识规则。因此，可以通过吸收市场实践共识，来逐步确定相关权属分配和竞争规则，建立数据市场基本秩序。例如在近年来的司法实践中，逐步建立了“用户授权”+“平台授权”+“用户授权”的三重授权原则来解决权属问题。这些司法认定与规则，与今年四月中央《关于构建更加完善的要素市场化配置体制机制的意见》》中指出的“研究根据数据性质完善产权性质”这一基本精神相符，在一定程度上，可以为未来数据权属的相关立法提供参考支撑。

七、展望与建议

党的十九届五中全会指出，要坚持创新在我国现代化建设全局中的核心地位，把科技自立自强作为国家发展的战略支撑，坚定不移建设制造强国、质量强国、网络强国、数字中国。在《中共中央关于制定国民经济和社会发展第十四个五年规划和二〇三五年远景目标的建议》中，明确提出要建立数据资源产权、交易流通、跨境传输和安全保护等基础制度和标准规范，推动数据资源开发利用，扩大基础公共信息数据有序开放，建设国家数据统一共享开放平台，保障国家数据安全，加强个人信息保护等。这对我国大数据的发展提出了更高、更深刻的要求。

作为生产要素，数据在国民经济运行中变得越来越重要，数据对经济发展、社会生活和国家治理已经在产生着根本性、全局性、革命性的影响。回望“十三五”时期，我国大数据发展依然面临着诸多问题。展望未来，我们期待大数据发展可以在下一个五年内继续突破。

一是进一步推动技术向更加贴近应用需求的方向发展。优化和提升基础类技术产品的架构和能力以降低应用成本；加强在自动化、智能化数据管理工具方面的研发投入；探索挖掘图分析技术的更多落地场景；扎实推进隐私计算的技术研发，在保障技术安全性的同时着力提升产品性能。

二是进一步提升数据应用的智能化、平民化。增强企业、政府、社会与个人的数据应用意识，打造数据驱动的社会文化模式；加强人才队伍建设，培养更多数据与业务双向精通的复合型人才，逐步促进大数据技术融合到日常业务和个人生活的各个环节。

三是进一步构建更加完善的数据治理体系。加速推进 DCMM 的贯标评估工作，在微观层面上增强组织内部的数据管理意识；优化数据共享、交易、流通相关的制度规则，打造可信数据服务体系，从宏观层面上构建自由有序的数据流通环境；同时，强化数据安全监管，开展企业级的数据安全治理评估工作，全面推进各行业各领域的数据安全治理能力建设。

四是进一步加强数据法律制度体系建设。在强化个人信息保护、数据跨境流动、数据权属等方面立法的同时，加强反垄断、反不正当竞争立法执法，通过考察市场的发展趋向、加强相关立法执法的国际合作与交流、提升执法团队的专业化水平等改善现行立法制度供给不足的现状，维护大数据行业发展秩序。

“明者见于无形，智者虑于未萌。”我国提出国家大数据战略，是紧跟时代步伐顺应历史规律的发展要求，是着眼全球提升国际综合竞争力的客观要求，是立足国情推动新旧动能接续转换的内在要求。以信息化培育新动能，用新动能推动新发展，以新发展创造新辉煌。

我们相信，以大数据为代表的新一代信息技术和产业的发展对于数字中国建设，乃至对于全面建设社会主义现代化国家的征程都将起到至关重要的推动作用。

参考文献

[1] 关于构建更加完善的要素市场化配置体制机制的意见[Z]. 中共中央，国务院，2020.
[2] 关于新时代加快完善社会主义市场经济体制的意见[Z]. 中共中央，国务院，2020.
[3] Digital Economy Compass 2020[R]. Statista.2020.
[4] 数据新视界[R].希捷，2020.
[5] 周涛，程学旗，陈宝权. CCF 大专委 2020 年大数据发展趋势预测. 大数据[J], 2020, 6(1): 119-123
[6] Top 10 Trends in Data and Analytics, 2020[R]. Gartner.2020.
[7] 2019 年软件和信息技术服务业统计年报[R]. 工信部运行局,2020.
[8] 李桥兴，胡雨晴. 大数据产业的属性与分类界定及其模糊识别研究[J]. 科技管理研究, 2020, 040(003):163-173.
[9] 中国大数据产业生态地图暨中国大数据产业发展白皮书[R]. 大数据产业生态联盟,2020.
[10] 中国智能风控市场调研报告[R]. 亿欧智库,2018
[11] 2018-2019 年度金融科技安全分析报告[R]. 普华永道，中国信息通信研究院，平安金融安全研究院
[12] 李伟. 做好数据治理推动数字化转型[J]. 中国金融, 2020, No.919(01):42-44.
[13] GB/T 36073-2018，数据管理能力成熟度评估模型[S].
[14] 2019-2020 年中国金融行业数据治理调研报告[R]. 中国计算机用户协会，2020.
[15] 工业企业数据资产管理现状调查报告[R]. 中国信息通信研究院，2018
[16] 2020 联合国电子政务调查报告[R]. 联合国，2020.
[17] 2020 中国地方政府数据开放报告[R]. 复旦大学联合国家信息中心数字中国研究院，2020.
[18] 陈炀. 我国大数据交易的风险问题及规则的构建[J]. 商情，2020，(2)：198.
[19] 北京国际大数据交易所设立工作实施方案[Z]. 北京市地方金融监督管理局，北京市经济和信息化局，2020.
[20] 深圳建设中国特色社会主义先行示范区综合改革试点实施方案（2020-2025 年）[Z]. 中共中央办公厅，国务院办公厅, 2020.
[21] GB/T 37988-2019, 信息安全技术 数据安全能力成熟度模型[S].
[22] 自治区厅局数据治理试点协同推进工作方案[Z]. 广西壮族自治区大数据发展局，2020.
[23] 郭小伟，孔硕. 个人信息权的规范构造——兼评三部《个人信息保护法》(专

家建议稿)[J]. 中国集体经济, 2020(26).
[24] 安宝双. 跨境数据流动:法律规制与中国方案[J]. 信息安全与技术, 2020, 011(003):1-6.
[25] 姜颖. 全球数据跨境流动发展态势及建议，中国信息通信研究院政策与经济研究所，2020。
[26] 詹馥静,王先林. 反垄断视角的大数据问题初探[J]. 价格理论与实践，2018，(9): 37.

中国信息通信研究院

地址：北京市海淀区花园北路 52 号

邮政编码：100191

联系电话：13683007576

传真：010-62304980

网址：www.caict.ac.cn

